ВЫСШАЯ МАТЕМАТИКА

специальные главы

• ВЫСШАЯ ШКОЛА •

П. И. ЧИНАЕВ, Н. А. МИНИН,
А. Ю. ПЕРЕВОЗНИКОВ, А. А. ЧЕРЕНКОВ

ВЫСШАЯ МАТЕМАТИКА

СПЕЦИАЛЬНЫЕ ГЛАВЫ

Под общей редакцией профессора П. И. Чинаева

ИЗДАНИЕ ВТОРОЕ

Киев
Головное издательство издательского объединения
«Вища школа»
1981

ББК 22.11
517
В93

Рекомендовано секцией математики и механики Научно-технического совета Минвуза УССР для использования студентами вузов

УДК 51 (07)

Высшая математика. Специальные главы : Пособие для студентов вузов / П. И. Чинаев, Н. А. Минин, А. Ю. Перевозников, А. А. Черенков. Под ред. П. И. Чинаева.— 2-е изд.— Киев : Вища школа. Головное изд-во, 1981.— 368 с.— 20203. 1702000000.

В пособии изложены следующие специальные главы высшей математики: основы теории функций комплексной переменной, интеграл Фурье, операционное исчисление, теория поля, уравнения математической физики, вариационное исчисление, основы теории матриц и линейной алгебры, понятие о линейном и динамическом программировании, приближенные вычисления. Приведено большое количество примеров и задач, в том числе и прикладного характера.

Предназначено для студентов вузов. Пособием могут пользоваться инженеры и лица, изучающие математику самостоятельно.

Табл. 19. Ил. 121. Список лит.: 12-назв.

Рецензенты: доц. Дубас В. И., ст. преподаватель Оскрого В. Г.

Редакция литературы по математике и физике
Зав. редакцией *Е. Л. Корженевич*

В $\frac{20203-035}{M211\ (04)-81}$ 102—81 1702000000

© Издательское объединение «Вища школа». 1977

© Издательское объединение «Вища школа». 1981

ОГЛАВЛЕНИЕ

Предисловие 3

Глава I. Основы теории функций комплексной переменной

§ 1. Комплексные числа. Области и границы 5
§ 2. Функции комплексной переменной 12
§ 3. Дифференцируемость и аналитичность функций комплексной переменной 14
§ 4. Интегрирование функций комплексной переменной. Определение интеграла, его основные свойства 19
§ 5. Ряды аналитических функций 26
§ 6. Вычеты 40
§ 7. Основные сведения о конформном отображении 48

Глава II. Интеграл Фурье

§ 8. Интеграл Фурье 52
§ 9. Некоторые частные случаи представления функции интегралом Фурье 56
§ 10. Комплексная форма интеграла Фурье. Преобразование Фурье 57
§ 11. Примеры разложения функций в интеграл Фурье 60

Глава III. Уравнения математической физики

§ 12. Понятие дифференциального уравнения с частными производными 67
§ 13. Классификация линейных дифференциальных уравнений с частными производными второго порядка 72
§ 14. Решение одномерного волнового уравнения 76
§ 15. Некоторые специальные функции 89
§ 16. Решение трехмерного однородного волнового уравнения 97

Глава IV. Операционное исчисление

§ 17. Преобразование Лапласа 103
§ 18. Основные теоремы операционного исчисления 110
§ 19. Некоторые приложения операционного исчисления 126
§ 20. Связь интеграла Лапласа с интегралом Фурье. Формула обращения 142

Глава V. Теория поля

§ 21. Понятие поля 143
§ 22. Скалярное поле. Градиент 144
§ 23. Векторное поле. Векторные линии поля 149
§ 24. Поток вектора 152
§ 25. Расходимость поля. Формула Остроградского 157
§ 26. Циркуляция вектора 161
§ 27. Вихрь (ротор) вектора. Формула Стокса 168
§ 28. Оператор Гамильтона и его применение 169
§ 29. Потенциальные поля 174
§ 30. Соленоидальные поля 182
§ 31. Основные операции векторного анализа в криволинейных координатах 187

Глава VI. Основы линейной алгебры и теории матриц

§ 32. Линейные функции и линейные преобразования 192
§ 33. Матрицы 193
§ 34. Линейные векторные пространства 208
§ 35. Системы линейных уравнений 227
§ 36. Примеры применения аппарата линейной алгебры и матричного исчисления 235

Глава VII. Вариационное исчисление

§ 37. Основные понятия и определения 239
§ 38. Экстремум функционала. Необходимое условие экстремума 245
§ 39. Условный экстремум функционала 259
§ 40. Вариационные задачи с подвижными границами, с угловыми точками, с ограничениями 265
§ 41. Понятие о достаточных условиях экстремума функционала 274
§ 42. Некоторые прямые методы вариационного исчисления 283
§ 43. Понятие о принципе максимума 286

Глава VIII. Линейное и динамическое программирование

§ 44. Задача линейного программирования 293
§ 45. Геометрическое представление задачи линейного программирования и ее решения 304
§ 46. Симплексный метод 313
§ 47. Понятие о динамическом программировании 324

Глава IX. Приближенные вычисления

§ 48. Приближенные числа. Погрешности 332
§ 49. Приближенное решение алгебраических и трансцендентных уравнений
§ 50. Конечные разности 341
§ 51. Интерполирование функций 344
§ 52. Приближенное интегрирование функций 350
§ 53. Приближенные способы решения дифференциальных уравнений . . . 353
§ 54. Приближенные методы расчета переходных процессов в нелинейных электрических цепях 355

Список литературы 363
Предметный указатель 364

ПРЕДИСЛОВИЕ

В пособии изложены специальные главы высшей математики, являющиеся теоретической основой таких курсов, как «Теоретические основы электротехники», «Теоретические основы радиотехники», «Теория автоматического управления» и др. Книга рассчитана на лиц, знакомых с основами аналитической геометрии и классического математического анализа — дифференциальным и интегральным исчислениями функций одной и многих переменных, обыкновенными дифференциальными уравнениями, числовыми и функциональными, главным образом степенными рядами и рядами Фурье. Пособие написано на основе многолетнего опыта работы авторов в технических вузах.

Авторы стремились в одной книге с единой точки зрения, последовательно и компактно изложить материал, подкрепляя его достаточным количеством решенных примеров и задач, в том числе и прикладного характера. Главы размещены так, чтобы знание материала предыдущих глав облегчало усвоение последующих и, по возможности, исключало дублирование. В то же время все главы книги относительно независимы друг от друга, поэтому структура пособия позволяет изучать некоторые темы в другой последовательности.

Авторы выражают глубокую благодарность академику АН УССР И. И. Ляшко, а также рецензентам — доценту кафедры прикладной математики Киевского поли-

технического института В. И. Дубас и старшему преподавателю кафедры высшей математики Одесского политехнического института В. Г. Оскрого за ряд ценных указаний, способствующих улучшению содержания книги.

Отзывы и пожелания просим направлять по адресу: *252054, Киев-54, ул. Гоголевская, 7, Головное издательство издательского объединения «Вища школа», редакция литературы по математике и физике.*

ОСНОВЫ ТЕОРИИ ФУНКЦИЙ КОМПЛЕКСНОЙ ПЕРЕМЕННОЙ

Теория функций комплексной переменной является логически стройной математической дисциплиной, позволяющей производить математические операции в области комплексных чисел. Она имеет огромное значение не только для математики (алгебры, дифференциальных уравнений, аналитической теории чисел) и различных прикладных математических дисциплин (аэро- и гидродинамики, небесной механики, теории упругости), но и широко используется при решении многих инженерных задач.

Теория функций комплексной переменной находит многочисленные приложения в электротехнике и радиотехнике, в частности в теории электро- и радиоцепей переменного тока как при их анализе, так и при синтезе, а также в теории электромагнитного поля.

§ 1. КОМПЛЕКСНЫЕ ЧИСЛА. ОБЛАСТИ И ГРАНИЦЫ

Основные определения. *Комплексным числом* $z = a + bi$ *называется пара действительных чисел* a *и* b, *взятых в определенном порядке:* $z = (a, b)$.

Если $b = 0$, то $z = a$ — действительное число; если $a = 0$, то $z = bi$ — чисто мнимое число. Величина a называется действительной частью комплексного числа, bi — мнимой частью, b — коэффициентом при мнимой части комплексного числа. Пишут: $a = \operatorname{Re} z$, $b = \operatorname{Im} z$. Эти обозначения происходят от французских слов rèel — «действительный» и imaginaire — «мнимый».

Два комплексных числа $z_1 = a_1 + b_1 i$ *и* $z_2 = a_2 + b_2 i$ *называются равными* ($z_1 = z_2$) *тогда и только тогда, когда одновременно* $a_1 = a_2$ *и* $b_1 = b_2$.

Над комплексными числами в алгебраической форме $z_1 = a_1 + b_1 i$ и $z_2 = a_2 + b_2 i$ определены действия:

сложение по формуле

$$z_1 + z_2 = (a_1 + b_1 i) + (a_2 + b_2 i) = (a_1 + a_2) + (b_1 + b_2) i \qquad (1)$$

и умножение по формуле

$$z_1 \cdot z_2 = (a_1 + b_1 i) \cdot (a_2 + b_2 i) = (a_1 a_2 - b_1 b_2) + (a_1 b_2 + b_1 a_2) i. \qquad (2)$$

Из такого определения действий следует, что $i^2 = -1$ (i — это символ, квадрат которого равен -1).

Комплексные числа $z = a + bi$ и $\bar{z} = a - bi$ называются комплексно-сопряженными. Легко проверить, что

$$z \cdot \bar{z} = a^2 + b^2.$$

Кроме сложения и умножения над комплексными числами $z_1 = a_1 + b_1 i$ и $z_2 = a_2 + b_2 i$ определены и обратные им действия: вычитание по формуле

$$z_1 - z_2 = (a_1 - a_2) + (b_1 - b_2) i \qquad (3)$$

и деление по формуле

$$\frac{z_1}{z_2} = \frac{a_1 + b_1 i}{a_2 + b_2 i} = \frac{(a_1 + b_1 i)(a_2 - b_2 i)}{a_2^2 + b_2^2} = \frac{(a_1 a_2 + b_1 b_2) + (a_2 b_1 - a_1 b_2) i}{a_2^2 + b_2^2}. \qquad (4)$$

Нетрудно заметить, что все эти действия выполняются над комплексными числами так же, как и над обычными алгебраическими выражениями, но с учетом того, что $i^2 = -1$ и $\frac{1}{i} = \frac{i}{i^2} = -i$.

Геометрическое представление и формы записи комплексного числа. Комплексные числа получают геометрическое представление на координатной плоскости в декартовой прямоугольной системе координат. При этом на оси абсцисс откладывается действительная (вещественная) часть, а на оси ординат — коэффициент при мнимой части. Координатную плоскость в этом случае называют комплексной плоскостью, а координатные оси соответственно действительной осью и мнимой осью.

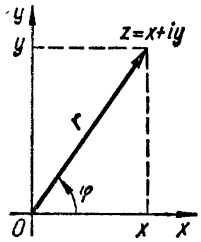

Рис. 1

Комплексное число z, записанное в виде $z = x + iy$, представляется на комплексной плоскости либо точкой (x, y), либо радиусом-вектором $z = \{x, y\}$. При этом длина радиуса-вектора $r = |z| = \sqrt{x^2 + y^2}$ называется модулем комплексного числа, а полярный угол φ радиуса-вектора, отсчитываемый от действительной оси против хода часовой стрелки (рис. 1),— аргументом комплексного числа ($\varphi = \operatorname{Arg} z$). Главное значение аргумента ($\arg z$) по определению выбирается в интервале $]-\pi; \pi]$, т. е. $-\pi < \arg z \leqslant \pi$.

Нетрудно установить, что условие $|z| = R$ ($R = \text{const}$) определяет окружность радиуса R с центром в начале координат (рис. 2, *а*); условие $|z| < R$ — открытый круг (без окружности) радиуса R с центром в начале координат (рис. 2, *б*); условие $|z - z_0| < R$ — открытый круг радиуса R с центром в точке z_0 (рис. 2, *в*); условие $\arg z = \text{const}$ — луч, исходящий из начала координат с полярным углом $\varphi = \arg z$ (рис. 2, *г*).

Если x и y — переменные величины, то комплексная величина $z = x + iy$ будет также переменной величиной. Тогда говорят, что z — комплексная переменная величина (коротко: комплексная переменная). Для комплексной переменной сохраняются понятия модуля $|z|$ и аргумента $\operatorname{Arg} z$ (также $\arg z$), однако, здесь эти величины будут переменными.

В соответствии с принятыми обозначениями (рис. 1) имеют место соотношения:

$$x = r\cos\varphi \quad \text{и} \quad y = r\sin\varphi, \tag{5}$$

откуда

$$z = r(\cos\varphi + i\sin\varphi). \tag{6}$$

Такая запись комплексного числа называется его тригонометрической формой.

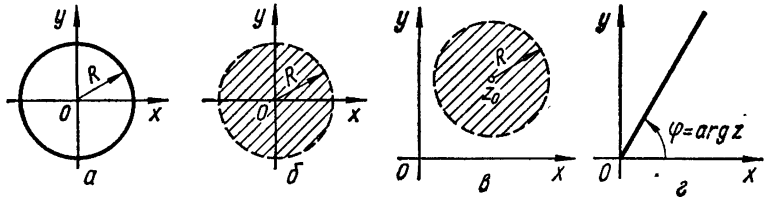

Рис. 2

Л. Эйлер доказал следующее тождество: $e^{i\varphi} = \cos\varphi + i\sin\varphi$. Используя это тождество, формулу (6) можно записать в виде

$$z = re^{i\varphi}. \tag{7}$$

Это показательная форма комплексного числа.

Для двух комплексных чисел $z_1 = r_1(\cos\varphi_1 + i\sin\varphi_1)$ и $z_2 = r_2(\cos\varphi_2 + i\sin\varphi_2)$ равенство возможно тогда и только тогда, когда $r_1 = r_2$, $\varphi_1 = \varphi_2 + 2\pi k$, где $k \in Z$ (вследствие периодичности тригонометрических функций).

Нетрудно видеть, что $|\bar{z}| = |z|$, но $\arg\bar{z} = -\arg z$.

Сложение и вычитание комплексных чисел в тригонометрической форме производятся так же, как и в алгебраической форме. Геометрически эти действия выполняются соответственно как сложение и вычитание векторов.

Умножение комплексных чисел в тригонометрической форме $z_1 = r_1(\cos\varphi_1 + i\sin\varphi_1)$ и $z_2 = r_2(\cos\varphi_2 + i\sin\varphi_2)$ осуществляется как обычно. Однако известные тригонометрические формулы позволяют упростить правило умножения:

$$z_1 \cdot z_2 = r_1(\cos\varphi_1 + i\sin\varphi_1) \cdot r_2(\cos\varphi_2 + i\sin\varphi_2) =$$
$$= r_1 \cdot r_2(\cos(\varphi_1 + \varphi_2) + i\sin(\varphi_1 + \varphi_2)) = r_1 \cdot r_2 e^{i(\varphi_1+\varphi_2)}$$

— при умножении двух комплексных чисел их модули перемножаются $|z_1 \cdot z_2| = |z_1| \cdot |z_2|$, *а аргументы складываются* $\operatorname{Arg}(z_1 \cdot z_2) = \operatorname{Arg} z_1 + \operatorname{Arg} z_2$ и $\arg(z_1 \cdot z_2) = \arg z_1 + \arg z_2$.

Геометрически умножение комплексного числа z_1 на комплексное число z_2 сводится к растяжению (сжатию) вектора z_1 в $|z_2|$ раз и повороту его на угол $\varphi_2 = \arg z_2$ (на рис. 3 треугольник $O1z_1$ подобен треугольнику Oz_2z).

Простыми следствиями правила умножения комплексных чисел являются:

правило возведения комплексного числа в целую положительную степень

$$z^n = (r(\cos\varphi + i\sin\varphi))^n = r^n(\cos n\varphi + i\sin n\varphi) = r^n e^{in\varphi}$$

и формула Муавра

$$(\cos\varphi + i\sin\varphi)^n = \cos n\varphi + i\sin n\varphi.$$

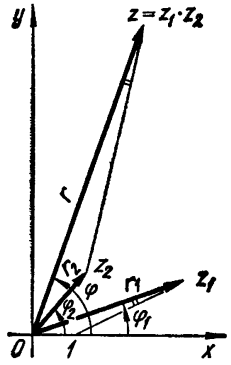

Рис. 3

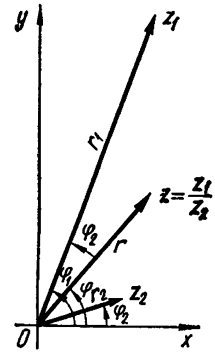

Рис. 4

Деление комплексных чисел в тригонометрической форме выполняется как действие, обратное умножению:

$$\frac{z_1}{z_2} = \frac{r_1}{r_2}(\cos(\varphi_1 - \varphi_2) + i\sin(\varphi_1 - \varphi_2)) = \frac{r_1}{r_2}e^{i(\varphi_1-\varphi_2)}$$

— *при делении комплексных чисел их модули делятся* $\left|\frac{z_1}{z_2}\right| = \frac{|z_1|}{|z_2|}$, *а аргументы вычитаются* $\operatorname{Arg}\frac{z_1}{z_2} = \operatorname{Arg} z_1 - \operatorname{Arg} z_2$ *и* $\arg\frac{z_1}{z_2} = \arg z_1 - \arg z_2$.

Геометрически деление комплексного числа z_1 на комплексное число z_2 сводится к сжатию (растяжению) вектора z_1 в $|z_2|$ раз и повороту его на угол $\varphi = \arg z_2$ по ходу часовой стрелки при $\varphi_2 > 0$ (рис. 4).

Извлечение корня степени n из комплексного числа z осуществляется как действие, обратное возведению в степень, с учетом периодичности тригонометрических функций:

$$\sqrt[n]{z} = \sqrt[n]{r(\cos\varphi + i\sin\varphi)} = \sqrt[n]{r}\left(\cos\frac{\varphi}{n} + i\sin\frac{\varphi}{n}\right) = \sqrt[n]{r}e^{i\frac{\varphi}{n}},$$

где $\varphi = \varphi_0 + 2\pi k$, $k \in Z_0$.

Таким образом,

$$\sqrt[n]{r(\cos\varphi + i\sin\varphi)} = \sqrt[n]{r}\left(\cos\frac{\varphi_0 + 2\pi k}{n} + i\sin\frac{\varphi_0 + 2\pi k}{n}\right) =$$

$$= \sqrt[n]{r}\, e^{i\frac{\varphi_0 + 2\pi k}{n}} = \sqrt[n]{r}\, e^{i\frac{\varphi_0}{n}} \cdot e^{i\frac{2\pi k}{n}},$$

$k = 0, 1, 2, \ldots, n-1$. Если $k = n$, то $\frac{\varphi}{n} = \frac{\varphi_0}{n} + 2\pi$, и получится то же значение корня, что и при $k = 0$. Так, между прочим, показывается, что корень степени n из комплексного числа z имеет точно n различных значений. Модули и аргументы всех этих значений равны соответственно $\sqrt[n]{|z|}$ и $\frac{\varphi_0}{n} + \frac{2\pi k}{n}$, т. е. все корни расположены на окружности радиуса $\sqrt[n]{r}$ в вершинах правильного n-угольника.

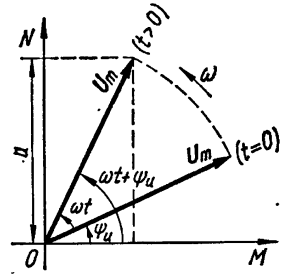

Рис. 5

В электротехнике при расчете цепей переменного тока широко применяется так называемый символический метод, который основан на использовании записи комплексного числа $a + ib$ в тригонометрической и показательной формах:

$$a + ib = A(\cos\alpha + i\sin\alpha) = Ae^{i\alpha},$$

где $A = \sqrt{a^2 + b^2}$, $\alpha = \operatorname{arctg}\frac{b}{a}$ $\left(-\frac{\pi}{2} < \alpha < \frac{\pi}{2}\right)$.

Если считать, что a и b — функции времени t, то геометрически выражение $a(t) + ib(t)$ будет представлять собой переменный радиус-вектор (при постоянном модуле — вращающийся радиус-вектор постоянной длины). Здесь рассматривается комплексная вектор-функция, изменяющаяся во времени по синусоидальному закону с угловой частотой ω:

$$a(t) + ib(t) = U_m(\cos(\omega t + \psi_u) + i\sin(\omega t + \psi_u)) = U_m e^{i\psi_u} \cdot e^{i\omega t}.$$

Начальное положение (при $t = 0$) этой вектор-функции характеризуется модулем U_m и аргументом ψ_u (рис. 5).

Вещественная часть $U_m\cos(\omega t + \psi_u)$ и коэффициент при мнимой части $U_m\sin(\omega t + \psi_u)$ являются соответственно проекциями на ось абсцисс и на ось ординат вектора длины U_m, вращающегося с постоянной угловой скоростью ω. Так как исследуется синусоидальная функция, то ограничиваются изучением только одной составляющей $u = U_m\sin(\omega t + \psi_u)$, которой ставится в соответствие вся вектор-функция, что записывается так:

$$u = U_m\sin(\omega t + \psi_u) \doteq U_m e^{i\psi_u} \cdot e^{i\omega t}$$

(рассматривать вторую составляющую нет необходимости, так как

$$\cos(\omega t + \psi_u) = \sin\left((\omega t + \psi_u) + \frac{\pi}{2}\right).$$

Множитель $U_m e^{i\psi_n}$, не зависящий от времени, обозначается $\dot{U}_m$ и называется комплексной амплитудой (в данном случае синусоидального напряжения u, имеющего начальную фазу ψ_u).

Таким образом,

$$u \doteq \dot{U}_m e^{i\omega t}.$$

Далее при расчетах цепей переменного тока оперируют только постоянными во времени комплексными числами типа $\dot{U}_m$ ($\dot{U}_m = U_m e^{i\psi_u}$), а не синусоидальными функциями времени как таковыми. После окончания расчета возвращаются к обычной форме записи синусоидальных функций.

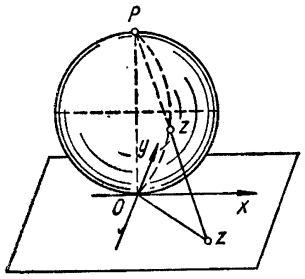

Рис. 6

Сущность символического метода состоит в переходе от операций над синусоидальными функциями времени к алгебраическим действиям над изображающими их постоянными комплексными числами и в возврате к синусоидальным функциям по окончании расчета.

Понятие о стереографической проекции. Выше было рассмотрено представление комплексных чисел на комплексной плоскости. Существует еще одно геометрическое представление комплексных чисел, с помощью которого вводится понятие бесконечно удаленной точки. Оно состоит в следующем. Поместим сферу произвольного радиуса так, чтобы она касалась комплексной плоскости в начале координат (рис. 6). Точку касания назовем южным полюсом сферы, а диаметрально противоположную точку — северным полюсом. Началу координат комплексной плоскости поставим в соответствие южный полюс сферы и припишем ему значение $z = 0$. Каждой другой точке z комплексной плоскости поставим в соответствие на сфере точку z пересечения со сферой луча, проведенного из северного полюса в точку z плоскости. Так будет установлено взаимно однозначное соответствие между комплексной плоскостью и сферой с выколотой точкой — северным полюсом.

Изображение комплексной плоскости на сфере называется стереографической проекцией комплексной плоскости, а сама сфера — числовой сферой.

Для того чтобы распространить взаимно однозначное соответствие комплексной плоскости на всю сферу, к комплексной плоскости присоединяют условную бесконечно удаленную точку ($z = \infty$), считая, что она соответствует северному полюсу. Эта точка ни в каких арифметических или алгебраических действиях не участвует,

однако к ней могут сходиться, и часто сходятся, последовательности комплексных чисел.

Комплексную плоскость с присоединенной к ней бесконечно удаленной точкой называют полной комплексной плоскостью, а без такой точки — открытой комплексной плоскостью.

Области и границы. Рассмотрим некоторые геометрические понятия.

Определение 1. *Множество D точек комплексной плоскости называется областью, если оно удовлетворяет следующим условиям:*

1) *вместе с каждой точкой из D этому множеству принадлежит и достаточно малый круг с центром в указанной точке* (свойство открытости);

2) *любые две точки множества D можно соединить ломаной, целиком состоящей из точек множества D* (свойство связности).

Определение 2. *Все точки, удовлетворяющие условию* 1) *определения области, называются точками области (внутренними точками области).*

Под внешней точкой области D понимают такую точку комплексной плоскости, которая не принадлежит множеству D вместе с некоторым кругом, содержащим эту точку.

Определение. 3. *Любая область, содержащая некоторую точку z_0, называется окрестностью этой точки. Обычно в качестве окрестности точки берут круг: если радиус круга r ($|z - z_0| < r$), то окрестность называется r-окрестностью точки z_0.*

Под окрестностью бесконечно удаленной точки понимают внешность любого круга с центром в начале координат (т. е. область $|z| > R$).

Определение 4. *Точка M называется граничной точкой области D, если она не принадлежит области D, но любая достаточно малая окрестность указанной точки содержит точки области D.*

Совокупность всех граничных точек области D называется ее границей. Например, для круга $|z| < 1$ границей является окружность $|z| = 1$.

Определение 5. *Всякая область D с присоединенной к ней границей называется замкнутой областью и обозначается через $\overline{D}$.*

Так, множество точек, удовлетворяющих неравенству $|z| \leqslant 2$, есть замкнутая область.

Для дальнейшего напомним следующие понятия из математического анализа:

Непрерывная на $[\alpha; \beta]$ функция $z(t)$ действительного параметра t определяет непрерывную кривую (линию, дугу); значения функции $z(t)$ называются точками кривой; уравнение $z = z(t)$ — это уравнение кривой (параметрическое).

На каждой кривой можно зафиксировать одно из двух направлений — в сторону возрастания параметра или, наоборот, в сторону убывания параметра. В первом случае $z(\alpha)$ — н а ч а л о кривой, а $z(\beta)$ — ее к о н е ц (во втором случае наоборот). *Если начало и конец кривой совпадают, то кривая называется з а м к н у т о й.*

Точка, соответствующая только одному значению параметра, называется п р о с т о й точкой; точка, соответствующая двум

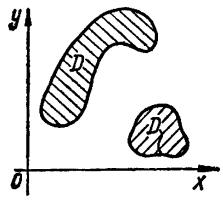

Рис. 7

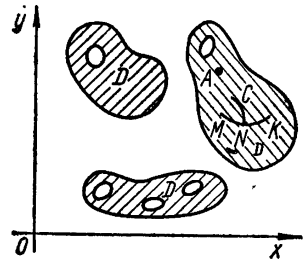

Рис. 8

и более значениям параметра, называется к р а т н о й точкой. Кривая, состоящая только из простых точек (не имеющая кратных точек), называется п р о с т о й (или ж о р д а н о в о й) кривой.

О п р е д е л е н и е 6. *Область D называется о д н о с в я з н о й, если любая простая замкнутая кривая, целиком принадлежащая D, может быть стянута в точку с помощью непрерывной деформации без выведения из области D.*

Из этого определения следует, что если область многосвязна, то ее граница не может состоять из одной простой замкнутой кривой.

Примеры односвязных областей приведены на рис. 7, многосвязных — на рис. 8.

В дальнейшем будут рассматриваться те случаи, когда граница области представляет собой одну или несколько кусочно-гладких кривых, которые, в частности, могут вырождаться в точки. В соответствии с этим введем два термина: «разрез» области (плоскости) внутри области или от ее границы и «выколотая точка».

§ 2. ФУНКЦИИ КОМПЛЕКСНОЙ ПЕРЕМЕННОЙ

Многие вопросы этого параграфа являются обобщением известных фактов из курса математического анализа функций действительной переменной.

Приведем здесь основные определения и понятия.

О п р е д е л е н и е 1. *На множестве M комплексной плоскости Z определена функция комплексной переменной $w = f(z)$, если каждой точке z множества M по определенному закону поставлены в соответствие одна или несколько точек w.* В первом случае функция о д н о з н а ч н а я, во втором — м н о г о з н а ч н а я.

Комплексная функция целочисленного аргумента называется последовательностью.

Для функции $w = f(z)$, как и для любой комплексной переменной, остаются в силе понятия модуля $|w|$ и аргумента $\operatorname{Arg} w$ (соответственно $\arg w$).

Если учесть, что $z = x + iy$, и положить $w = u + iv$, то станет ясным, что для определения функции w достаточно определить две функции действительных переменных: $u = u(x, y)$ и $v = v(x, y)$.

Геометрическое представление функции $w = f(z)$ можно получить, вводя вторую комплексную плоскость W (uO_1v). При этом переменная z будет представлена точками на плоскости xOy, а переменная w — точками на плоскости uO_1v. Геометрическое представление множества M на плоскости Z будет образом, а соответствующее ему представление на плоскости W множества точек w (назовем его множеством N) — отображением.

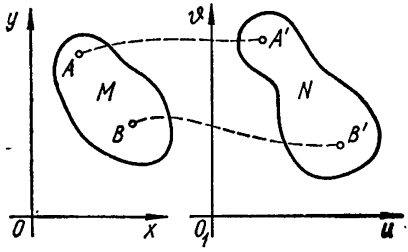

Рис. 9

Пусть функция $w = f(z)$ отображает множество M на множество N. Тогда функция $z = f^{-1}(w)$, ставящая в соответствие точке w из множества N совокупность всех точек z из множества M (которые заданная функция $w = f(z)$ отображает в точки w), называется обратной функцией по отношению к функции $f(z)$. Особенно важную роль в дальнейшем играет случай, когда функция $w = f(z)$ однозначна в M, а обратная ей функция $z = f^{-1}(w)$ однозначна в N; в этом случае отображение $w = f(z)$ взаимно однозначно в M. При таком отображении двум различным точкам в M всегда соответствуют две различные точки в N и, наоборот, двум различным точкам в N всегда соответствуют две различные точки в M (рис. 9).

Если функция $w = f(z)$ отображает множество M на множество N, а функция $\omega = g(w)$ — множество N на множество P, то функция $\omega = h(z) = g(f(z))$ называется сложной функцией от z, составленной из f и g, а отображение h множества M на множество P — суперпозицией отображений f и g.

Введем понятие предела функции комплексной переменной.

Определение 2. *Комплексное число $w_0 = u_0 + iv_0$ называется пределом функции $w = u(x, y) + iv(x, y) = f(z)$ комплексной переменной $z = x + iy$ при $z \to z_0$, если для любого, сколь угодно малого наперед заданного положительного числа ε можно указать такое положительное число δ, что из неравенства $|z - z_0| < \delta$ следует неравенство $|f(z) - w_0| < \varepsilon$.*

Это записывается так: $\lim\limits_{z \to z_0} f(z) = w_0$.

Из определения следует, что если значение w_0 есть предел функции $w = f(z)$ при $z \to z_0$, то это значение w_0 не зависит от пути,

по которому z приближается к z_0. Из определения следует также, что если этот предел существует, то существуют и пределы

$$\lim_{\substack{x\to x_0\\ y\to y_0}} u(x, y) = u_0 \quad \text{и} \quad \lim_{\substack{x\to x_0\\ y\to y_0}} v(x, y) = v_0.$$

Для функции комплексной переменной справедливы такие же теоремы о пределах, как и для функции действительной переменной. В частности, если пределы существуют, то

$$\lim_{z\to z_0} (c_1 f_1(z) \pm c_2 f_2(z)) = c_1 \lim_{z\to z_0} f_1(z) \pm c_2 \lim_{z\to z_0} f_2(z),$$

где c_1 и c_2 — комплексные постоянные,

$$\lim_{z\to z_0} (f_1(z) \cdot f_2(z)) = \lim_{z\to z_0} f_1(z) \cdot \lim_{z\to z_0} f_2(z)$$

и

$$\lim_{z\to z_0} \frac{f_1(z)}{f_2(z)} = \frac{\lim\limits_{z\to z_0} f_1(z)}{\lim\limits_{z\to z_0} f_2(z)}, \quad \text{если} \quad \lim_{z\to z_0} f_2(z) \neq 0.$$

Определение 3. *Функция $f(z)$ называется непрерывной функцией в точке z_0, если она определена в некоторой окрестности этой точки и* $\lim\limits_{z\to z_0} f(z) = f(z_0)$.

Функция $f(z)$ непрерывна в области D, если она непрерывна в каждой точке этой области.

Непрерывные функции комплексной переменной обладают теми же свойствами, что и непрерывные функции действительной переменной.

В частности — *функция $w = f(z)$, непрерывная в замкнутой области $\overline{D}$:*

1) *ограничена по модулю,*

2) *достигает своего наименьшего и наибольшего значений по модулю,*

3) *равномерно непрерывна в области $\overline{D}$.*

Кроме того, существует следующее важное свойство непрерывных функций комплексной переменной [1]:

Если функция $w = f(z)$ непрерывна в области $\overline{D}$ и реализует взаимно однозначное отображение этой области на некоторое множество Δ в плоскости W, то Δ также является областью, и обратная функция $z = f^{-1}(w)$ непрерывна в области Δ.

§ 3. ДИФФЕРЕНЦИРУЕМОСТЬ И АНАЛИТИЧНОСТЬ ФУНКЦИЙ КОМПЛЕКСНОЙ ПЕРЕМЕННОЙ

Определение 1. *Пусть функция $f(z)$ определена в некоторой окрестности точки z. Тогда предел*

$$\lim_{\Delta z\to 0} \frac{f(z+\Delta z) - f(z)}{\Delta z} = f'(z),$$

[1] См.: М. А. Лаврентьев, Б. В. Шабат. Методы теории функций комплексного переменного. М., Наука, 1973, с. 21.

если он существует, называется производной от функции $f(z)$ *по комплексной переменной* z.

Для существования производной в точке z необходима непрерывность функции в этой точке.

Определение 2. *Функция дифференцируема в точке* z, *если ее приращение в этой точке* $\Delta f = f(z+\Delta z) - f(z)$ *может быть представлено в виде суммы главной части, линейной относительно* Δz, *и бесконечно малых высших порядков, т. е. если*

$$f(z+\Delta z) - f(z) = C \cdot \Delta z + \gamma(\Delta z) \cdot \Delta z,$$

причем $\gamma(\Delta z) \to 0$ при $\Delta z \to 0$.

Из определения следует, что

$$\frac{f(z+\Delta z) - f(z)}{\Delta z} = C + \gamma(\Delta z) \to C \quad \text{при} \quad \Delta z \to 0.$$

Это значит, что дифференцируемая в точке z функция имеет в этой точке производную $f'(z) = C$.

Легко доказывается и обратное утверждение: *если функция* $f(z)$ *имеет в некоторой точке производную, то она дифференцируема в этой точке.*

Теорема. *Если функция* $f(z) = u(x, y) + iv(x, y)$ *определена в некоторой окрестности точки* $z = x + iy$, *причем функции* $u(x, y)$ *и* $v(x, y)$ *в этой точке дифференцируемы, то для дифференцируемости* $f(z)$ *в точке* z *необходимо и достаточно, чтобы в этой точке имели место равенства*

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

Эти равенства называются условиями д'Аламбера — Эйлера.

Доказательство. *Необходимость.* Предположим, что функция $f(z)$ дифференцируема в точке z, т. е. существует

$$f'(z) = \lim_{\Delta z \to 0} \frac{f(z+\Delta z) - f(z)}{\Delta z}.$$

Поскольку предел не зависит от пути, по которому $\Delta z \to 0$, то предположим сначала, что Δz действительное, т. е. $\Delta z = \Delta x$, $\Delta y = 0$. Тогда

$$f'(z) = \lim_{\Delta x \to 0} \frac{(u(x+\Delta x, y) + iv(x+\Delta x, y)) - (u(x, y) + iv(x, y))}{\Delta x} =$$

$$= \lim_{\Delta x \to 0} \frac{(u(x+\Delta x, y) - u(x, y)) + i(v(x+\Delta x, y) - v(x, y))}{\Delta x} =$$

$$= \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}.$$

Теперь предположим, что Δz — чисто мнимое, т. е. $\Delta z = i\Delta y$, $\Delta x = 0$. В этом случае

$$f'(z) = \lim_{\Delta y \to 0} \frac{(u(x, y+\Delta y) + iv(x, y+\Delta y)) - (u(x, y) + iv(x, y))}{i\Delta y} =$$

$$= \lim_{\Delta y \to 0} \frac{(u(x, y+\Delta y) - u(x, y)) + i(v(x, y+\Delta y) - v(x, y))}{i\Delta y} =$$

$$= -i\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y} = \frac{\partial v}{\partial y} - i\frac{\partial u}{\partial y}.$$

Сравнивая вычисленные пределы, получаем

$$\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} - i\frac{\partial u}{\partial y},$$

откуда

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad \text{и} \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

Достаточность. Пусть теперь в некоторой окрестности точки z выполнены условия д'Аламбера — Эйлера

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}; \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

Докажем, что функция $f(z)$ дифференцируема.

Так как по условию теоремы функции $u(x, y)$ и $v(x, y)$ дифференцируемы, то их полные приращения могут быть представлены в виде

$$\Delta u = \frac{\partial u}{\partial x}\Delta x + \frac{\partial u}{\partial y}\Delta y + \alpha_1 |\Delta z|,$$

$$\Delta v = \frac{\partial v}{\partial x}\Delta x + \frac{\partial v}{\partial y}\Delta y + \alpha_2 |\Delta z|,$$

где $|\Delta z| = \sqrt{(\Delta x)^2 + (\Delta y)^2}$ и $\alpha_1 \to 0$, $\alpha_2 \to 0$ при $|\Delta z| \to 0$, и так как $\Delta f = \Delta u + i\Delta v$, $\Delta z = \Delta x + i\Delta y$, то

$$\frac{\Delta f}{\Delta z} = \frac{\Delta u + i\Delta v}{\Delta x + i\Delta y} =$$

$$= \frac{\frac{\partial u}{\partial x}\Delta x + \frac{\partial u}{\partial y}\Delta y + \alpha_1 |\Delta z| + i\left(\frac{\partial v}{\partial x}\Delta x + \frac{\partial v}{\partial y}\Delta y + \alpha_2 |\Delta z|\right)}{\Delta x + i\Delta y} =$$

$$= \frac{\left(\frac{\partial u}{\partial x}\Delta x + i\frac{\partial v}{\partial x}\Delta x\right) + \left(\frac{\partial u}{\partial y}\Delta y + i\frac{\partial v}{\partial y}\Delta y\right)}{\Delta x + i\Delta y} + \frac{(\alpha_1 + i\alpha_2)\cdot|\Delta z|}{\Delta z} =$$

$$= \frac{\left(\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}\right)\Delta x + i\left(i\frac{\partial v}{\partial x} + \frac{\partial u}{\partial x}\right)\Delta y}{\Delta x + i\Delta y} + (\alpha_1 + i\alpha_2)\frac{|\Delta z|}{\Delta z} =$$

$$= \frac{\left(\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}\right)(\Delta x + i\Delta y)}{\Delta x + i\Delta y} + (\alpha_1 + i\alpha_2)\frac{|\Delta z|}{\Delta z} =$$

$$= \left(\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}\right) + (\alpha_1 + i\alpha_2)\frac{|\Delta z|}{\Delta z}.$$

Отсюда следует, что $\lim\limits_{\Delta z \to 0} \frac{\Delta f}{\Delta z} = f'(z)$ существует. При этом $f'(z) = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}$. Теорема доказана.

Учитывая условия д'Аламбера — Эйлера, производную можно записать в одной из следующих четырех форм:

$$f'(z) = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}, \quad f'(z) = \frac{\partial v}{\partial y} + i\frac{\partial v}{\partial x},$$

$$f'(z) = \frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y}, \quad f'(z) = \frac{\partial v}{\partial y} - i\frac{\partial u}{\partial y}.$$

Поскольку свойства алгебраических действий и правила предельного перехода для функций действительной переменной распространяются и на функцию комплексной переменной, то правила дифференцирования функций действительной переменной справедливы и для функций комплексной переменной. В частности,

$$(f(z) + \varphi(z))' = f'(z) + \varphi'(z),$$

$$(f(z) \cdot \varphi(z))' = f'(z) \cdot \varphi(z) + f(z) \cdot \varphi'(z),$$

$$\left(\frac{f(z)}{\varphi(z)}\right)' = \frac{f'(z) \cdot \varphi(z) - f(z) \cdot \varphi'(z)}{\varphi^2(z)} \quad (\varphi(z) \neq 0),$$

$$(f(\varphi(z)))' = f'(\varphi) \cdot \varphi'(z),$$

$$f'(z) = \frac{1}{(f^{-1}(w))'},$$

где $f^{-1}(w)$ — функция, обратная функции $f(z)$ (при этом предполагается, что функции $f(z)$ и $f^{-1}(w)$ взаимно однозначны) и $(f^{-1}(w))' \neq 0$.

Определение. *Однозначная функция $f(z)$, дифференцируемая в каждой точке области D, называется аналитической (регулярной, моногенной) функцией в этой области.*

Замечания

1. Всякая функция $\varphi(x, y)$, удовлетворяющая в некоторой области D дифференциальному уравнению Лапласа (см. гл. III) $\frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} = 0$, называется гармонической функцией в этой области.

Нетрудно установить, что если $f(z) = u(x, y) + iv(x, y)$ — дифференцируемая функция, то функции $u(x, y)$ и $v(x, y)$ являются гармоническими функциями. Ниже будет доказано существование производных высших порядков для этих функций, а здесь пока убедимся в формальной справедливости высказанного утверждения.

Дифференцируя первое из уравнений д'Аламбера — Эйлера по y, а второе по x, находим

$$\frac{\partial^2 u}{\partial x \partial y} = \frac{\partial^2 v}{\partial y^2}, \quad \frac{\partial^2 v}{\partial x^2} = -\frac{\partial^2 u}{\partial x \partial y}, \text{ откуда } \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0.$$

2. Уравнения д'Аламбера — Эйлера позволяют, зная одну из функций u или v, определить с точностью до произвольной постоянной и вторую функцию, а следовательно, построить всю функцию $f(z) = u(x, y) + iv(x, y)$.

Примеры

1. Установить, что функция $e^z = u(x, y) + iv(x, y)$ является аналитической функцией, и найти ее производную.

Р е ш е н и е. Определяем функцию e^z, задавая ее действительную и мнимую части: $u(x, y) = e^x \cos y$, $v(x, y) = e^x \sin y$, т. е.

$$e^z = e^x \cos y + ie^x \sin y.$$

Покажем, что функция e^z имеет производную на всей комплексной плоскости Z. Проверим справедливость для e^z условий д'Аламбера — Эйлера. Имеем:

$$\frac{\partial u}{\partial x} = e^x \cos y, \quad \frac{\partial u}{\partial y} = -e^x \sin y;$$

$$\frac{\partial v}{\partial x} = e^x \sin y, \quad \frac{\partial v}{\partial y} = e^x \cos y.$$

Легко видеть, что для функции e^z выполняются условия д'Аламбера — Эйлера

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

Производная функции $f(z) = e^z = e^x \cos y + ie^x \sin y$ по z

$$f'(z) = (e^z)' = \frac{\partial y}{\partial x} + i\frac{\partial v}{\partial x} = e^x \cos y + ie^x \sin y = e^z.$$

2. Найти $f(z) = u(x, y) + iv(x, y)$, если

$$v(x, y) = 2(\operatorname{ch} x \sin y - xy), \quad f(0) = 0.$$

Р е ш е н и е. Учитывая уравнения д'Аламбера — Эйлера, найдем $\frac{\partial v}{\partial y} = 2 \operatorname{ch} x \cos y - 2x = \frac{\partial u}{\partial x}$. Отсюда

$$u(x, y) = \int \frac{\partial v}{\partial y}\, dx + \varphi(y) = \int (2 \operatorname{ch} x \cos y - 2x)\, dx + \varphi(y) =$$

$$= 2 \operatorname{sh} x \cos y - x^2 + \varphi(y).$$

Теперь удовлетворим второму уравнению д'Аламбера — Эйлера. Имеем

$$\frac{\partial v}{\partial x} = 2 \operatorname{sh} x \sin y - 2y,$$

$$\frac{\partial u}{\partial y} = -2 \operatorname{sh} x \sin y + \varphi'(y),$$

и так как $\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$, то $\varphi'(y) = 2y$, откуда

$$\varphi(y) = y^2 + C.$$

Поэтому

$$u(x, y) = 2 \operatorname{sh} x \cos y - x^2 + y^2 + C.$$

Строим функцию $f(z) = u(x, y) + iv(x, y)$:

$$f(z) = (2 \operatorname{sh} x \cos y - x^2 + y^2 + C) + i(2 \operatorname{ch} x \sin y - 2xy) =$$

$$= (e^x - e^{-x}) \cos y + i(e^x + e^{-x}) \sin y - (x^2 + 2ixy - y^2 + C) =$$

$$= (\cos y + i \sin y)\, e^{x} - (\cos y - i \sin y)\, e^{-x} - [x^2 + 2x\,(iy) + (iy)^2] + C =$$

$$= e^{x} e^{iy} - e^{-x} e^{-iy} - (x + iy)^2 + C = e^{x+iy} - e^{-(x+iy)} - (x + iy)^2 + C.$$

Учитывая, что $x + iy = z$, находим

$$f(z) = e^{z} - e^{-z} - z^2 + C,$$

и так как $f(0) = 0$, то $C = 0$.

Следовательно, $f(z) = e^{z} - e^{-z} - z^2$.

§ 4. ИНТЕГРИРОВАНИЕ ФУНКЦИЙ КОМПЛЕКСНОЙ ПЕРЕМЕННОЙ. ОПРЕДЕЛЕНИЕ ИНТЕГРАЛА, ЕГО ОСНОВНЫЕ СВОЙСТВА

Пусть на плоскости Z задана ориентированная [1] кривая C и на ней функция $f(z)$.

Разобьем кривую C на n произвольных частей $\Delta z_k = z_{k+1} - z_k$ (рис. 10). В каждой части выберем произвольную точку ζ_k, вычислим в ней значение функции $f(\zeta_k)$ и составим интегральную сумму $\sum_{k=1}^{n} f(\zeta_k)\, \Delta z_k$.

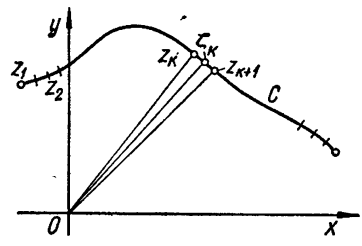

Рис. 10

Предел этой интегральной суммы при $\max |\Delta z_k| \to 0$ *(если он существует и не зависит от способа разбиения и выбора точек внутри каждой элементарной части) называется интегралом от функции* $f(z)$ *вдоль кривой (контура)* C *и обозначается* $\int_C f(z)\, dz$.

Если $f(z)$ — кусочно-непрерывная функция, а кривая C — кусочно-гладкая кривая, то интеграл $\int_C f(z)\, dz$ всегда существует. Это доказывается так же, как и в математическом анализе существование криволинейного интеграла от функции действительной переменной.

Если положить $f(z) = u(x, y) + iv(x, y)$, где $u(x, y)$ и $v(x, y)$ — функции действительных переменных, то интегралу $\int_C f(z)\, dz$ можно придать другую форму. Очевидно, $z_k = x_k + iy_k$, $x_{k+1} - x_k = \Delta x_k$, $y_{k+1} - y_k = \Delta y_k$

$$\zeta_k = \xi_k + i\eta_k,\ u(\xi_k, \eta_k) = u_k,\ v(\xi_k, \eta_k) = v_k.$$

Поэтому

$$\sum_{k=1}^{n} f(\zeta_k)\, \Delta z_k = \sum_{k=1}^{n} (u_k + iv_k)(\Delta x_k + i \Delta y_k) =$$

[1] Ориентированная кривая — это кривая, на которой указано положительное направление. Если кривая замкнута и не имеет точек самопересечения, то положительным будем считать такое направление, при движении в котором внутренние точки области, ограниченной этой кривой, остаются слева.

$$= \sum_{k=1}^{n} ((u_k \Delta x_k - v_k \Delta y_k) + i (v_k \Delta x_k + u_k \Delta y_k)) =$$

$$= \sum_{k=1}^{n} (u_k \Delta x_k - v_k \Delta y_k) + i \sum_{k=1}^{n} (u_k \Delta y_k + v_k \Delta x_k).$$

После перехода к пределу при $\max | \Delta z_k | \to 0$ получаем

$$\int_C f(z)\, dz = \int_C u dx - v dy + i \int_C u dy + v dx.$$

Таким образом, вычисление интеграла от функции комплексной переменной сводится к вычислению криволинейных интегралов от действительных функций действительных переменных.

Используя введенные определения и полученную формулу, можно сделать еще ряд важных выводов.

Прежде всего отметим, что если $w(t) = x(t) + iy(t)$ есть комплексная функция действительной переменной, то ее производная и интеграл вычисляются, как соответствующие линейные комбинации производных и интегралов от составляющих функций:

$$w'(t) = x'(t) + iy'(t),$$

$$\int_{t_1}^{t_2} w(t)\, dt = \int_{t_1}^{t_2} x(t)\, dt + i \int_{t_1}^{t_2} y(t)\, dt.$$

Далее, если кривая C задана параметрически $z = z(t)$, причем $t_1 < t < t_2$, то

$$\int_C f(z)\, dz = \int_{t_1}^{t_2} f(z(t))\, z'(t)\, dt.$$

Наконец, следует подчеркнуть, что для интегралов от комплексных функций комплексной переменной справедливы все свойства криволинейных интегралов от действительных функций действительных переменных.

Так, в частности,

1) $\int_C (A_1 f_1(z) + A_2 f_2(z))\, dz = A_1 \int_G f_1(z)\, dz + A_2 \int_G f_2(z)\, dz;$

2) $\int_C f(z)\, dz = \int_{C_1} f(z)\, dz + \int_{C_2} f(z)\, dz;$

3) $\int_{C^-} f(z)\, dz = -\int_C f(z)\, dz;$

4) *если на кривой* C *всюду* $| f(z) | \leqslant M$ $(M > 0)$, *то*

$$\left| \int_G f(z)\, dz \right| \leqslant Ml.$$

(В записанных выше выражениях A_1 и A_2 — комплексные постоянные; C_1 и C_2 — две кривые, составляющие кривую C; C^- — кривая,

совпадающая с кривой C по форме и положению, но с противоположным направлением обхода; l — длина дуги кривой C).

Помимо упомянутых свойств для аналитических функций справедлива следующая теорема.

Теорема Коши. *Если функция $f(z)$ аналитична в односвязной области D, то интеграл $\int_C f(z)\,dz$ не зависит от формы кривой C, а зависит только от положения конечных точек дуги.*

Доказательство. Для простоты доказательство теоремы проведем в предположении непрерывности производной $f'(z)$ (аналитичность функции предусматривает только существование $f'(z)$, но не ее непрерывность).

Пусть $f(z) = u(x, y) + iv(x, y)$. Тогда из непрерывности производной $f'(z)$ следует непрерывность частных производных $\frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial y}$, $\frac{\partial v}{\partial x}$, $\frac{\partial v}{\partial y}$. Известно, что $\int_C f(z)\,dz = \int_C u\,dx - v\,dy + i\int_C u\,dy + v\,dx$.

Этот интеграл не зависит от формы пути интегрирования, если от нее не зависят интегралы, стоящие в правой части равенства, а так как это криволинейные интегралы от действительных функций действительных переменных, то для их независимости от формы пути интегрирования необходимо и достаточно, чтобы

$$\frac{\partial(-v)}{\partial x} = \frac{\partial u}{\partial y} \quad \text{и} \quad \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}.$$

Функция $f(z)$ аналитична, поэтому указанные равенства имеют место (это уравнения д'Аламбера — Эйлера). Следовательно, интеграл $\int_C f(z)\,dz$ не зависит от формы пути интегрирования. Теорема доказана.

Из теоремы Коши непосредственно следует, что если $f(z)$ аналитична в односвязной области D, то функция

$$F(z) = \int_{z_0}^{z} f(\zeta)\,d\zeta$$

тоже аналитична в области D, причем,

$$F'(z) = \frac{d}{dz}\int_{z_0}^{z} f(\zeta)\,d\zeta = f(z).$$

Используя этот результат, можно показать, что если $F(z)$ — произвольная первообразная для функции $f(z)$, то

$$\int_{z_0}^{z} f(\zeta)\,d\zeta = F(z) - F(z_0),$$

а также, что $\int_C f(z)\,dz = \int_{z_1}^{z_2} f(z)\,dz$, если z_1 и z_2 — концы кривой C.

Из теоремы Коши следует также утверждение: *в односвязной области интеграл от аналитической функции, взятый по любому замкнутому контуру, равен нулю.*

Действительно, всякий замкнутый простой контур C всегда можно с помощью двух точек z_1 и z_2 разбить на два составляющих открытых контура C_1 и C_2 (рис. 11). Тогда

$$\oint_C f(z)\,dz = \int_{C_1} f(z)\,dz + \int_{C_2^-} f(z)\,dz = \int_{C_1} f(z)\,dz - \int_{C_2^+} f(z)\,dz = 0,$$

так как $\int_{C_1} f(z)\,dz = \int_{C_2} f(z)\,dz$, поскольку концы кривых C_1 и C_2 совпадают. (Символом $\oint$ часто обозначают интеграл по замкнутому контуру.)

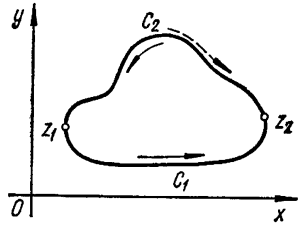

Рис. 11

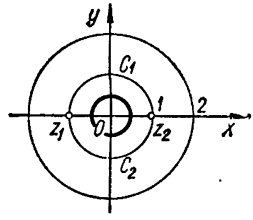

Рис. 12

Заметим, что в односвязной области справедлива теорема, обратная теореме Коши. В многосвязной области теорема Коши в сформулированном выше виде, вообще говоря, не имеет места. Это подтверждает хотя бы следующий пример.

Функция $f(z) = \frac{1}{z}$ аналитична в кольце $\frac{1}{2} < |z| < 2$ (рис. 12).

Возьмем в этой двусвязной области замкнутый контур C — окружность $|z| = 1$ и на нем две точки $z_1 = -1$ и $z_2 = 1$. Вычислим интегралы по кривым C_1 и C_2, соединяющим точки z_1 и z_2, где C_1 — верхняя, C_2 — нижняя полуокружности.

На окружности C имеем $|z| = 1$, $z = e^{i\varphi}$, $dz = ie^{i\varphi}\,d\varphi$. Поэтому

$$\int_{C_1} \frac{dz}{z} = \int_0^{\pi} \frac{ie^{i\varphi}d\varphi}{e^{i\varphi}} = \int_{\pi}^{0} id\varphi = i\varphi\Big|_{\pi}^{0} = -i\pi,$$

$$\int_{C_2} \frac{dz}{z} = \int_{-\pi}^{0} \frac{ie^{i\varphi}d\varphi}{e^{i\varphi}} = \int_{-\pi}^{0} id\varphi = i\varphi\Big|_{-\pi}^{0} = i\pi;$$

$$\int_{C_1} \frac{dz}{z} \neq \int_{C_2} \frac{dz}{z}.$$

Откуда

$$\oint_G \frac{dz}{z} \neq 0.$$

Для многосвязной области D справедливо утверждение:

Если функция $f(z)$ аналитична и замкнутый контур C охватывает односвязную часть D_1 внутри области D, то

$$\oint_C f(z)\,dz = 0.$$

В связи с этим и для многосвязной области можно сформулировать теорему Коши.

Теорема. *Пусть функция $f(z)$ аналитична в многосвязной области D и непрерывна в замкнутой многосвязной области $\overline{D}$ и пусть кривая C — составная граница этой области. Тогда, если при интегрировании вдоль C эту границу проходить так, что область D будет всегда расположена по одну сторону, то интеграл от функции $f(z)$ вдоль границы C будет равен нулю.*

Интеграл Коши. Формула Коши. Пусть D — односвязная область, ограниченная кусочно-гладкой линией C, $f(z)$ — функция, аналитическая в замкнутой области $\overline{D}$. Тогда имеет место следующая формула Коши:

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f(\zeta)\,d\zeta}{\zeta - z},$$

где z — любая точка внутри C, а интегрирование по контуру C совершается в положительном направлении.

Эта формула может быть распространена и на случай многосвязной области D, границей которой является сложный контур Γ, состоящий из конечного числа кусочно-гладких замкнутых линий.

Если $f(z)$ — аналитическая функция в замкнутой области $\overline{D}$, то

$$f(z) = \frac{1}{2\pi i}\int_\Gamma \frac{f(\zeta)\,d\zeta}{\zeta - z},$$

где z — любая точка области D, а интегрирование выполняется вдоль сложного контура Γ в положительном направлении.

Выражение

$$\frac{1}{2\pi i}\int_\Gamma \frac{f(\zeta)\,d\zeta}{\zeta - z},$$

где $f(z)$ — функция, аналитическая в замкнутой области $\overline{D}$, границей которой служит контур Γ, называется интегралом Коши.

Интеграл Коши изображает данную функцию $f(z)$, аналитическую в $\overline{D}$, во всякой точке, внутренней к контуру Γ.

Во всякой точке z, лежащей вне замкнутой области $\overline{D}$, интеграл Коши равен нулю, ибо функция $\frac{f(\zeta)}{\zeta - z}$ для всех точек z, не лежащих в области $\overline{D}$,— аналитическая.

Пусть L — произвольная кусочно-гладкая линия (не обязательно замкнутая), $\varphi(z)$ — заданная на L непрерывная функция.

Выражение

$$\frac{1}{2\pi i}\int_L \frac{\varphi(\zeta)\,d\zeta}{\zeta - z}$$

имеет определенное значение для любой точки z, не лежащей на L, и определяет однозначную функцию $F(z)$ во всех точках z, не принадлежащих L.

Это выражение называется интегралом типа Коши. Если L — замкнутая линия, а $\varphi(z)$ — функция, аналитическая внутри L и на L, получим интеграл Коши.

Теорема. *Функция $F(z)$, определенная интегралом типа Коши*

$$F(z) = \frac{1}{2\pi i}\int_L \frac{\varphi(\zeta)\,d\zeta}{\zeta - z},$$

есть аналитическая во всякой односвязной области D, не содержащей точек линии L, и для ее производной имеет место формула

$$F'(z) = \frac{1}{2\pi i}\int_L \frac{\varphi(\zeta)\,d\zeta}{(\zeta - z)^2}.$$

Из этой формулы видно, что для нахождения производной функции $F(z)$ нужно выполнить дифференцирование по параметру z в интеграле типа Коши.

Доказывается, что это дифференцирование можно повторить n раз, поэтому:

Функция $F(z)$, определяемая интегралом типа Коши, имеет в каждой точке z, не лежащей на L, производные всех порядков, причем

$$F^{(n)}(z) = \frac{n!}{2\pi i}\int_L \frac{\varphi(\zeta)\,d\zeta}{(\zeta - z)^{n+1}}.$$

Существование производных любого порядка для аналитической функции. Известно, что в случае функции действительной переменной из факта существования первой производной ничего нельзя сказать о существовании производных высших порядков. Для функций комплексной переменной справедлива следующая теорема:

Теорема. *Если однозначная функция $f(z)$ имеет всюду в области D первую производную, то она имеет в этой области производные любого порядка.*

Доказательство. Пусть z — произвольная точка области D, C — кусочно-гладкий замкнутый контур, окружающий точку

z и лежащий со всеми своими внутренними точками в области D. По формуле Коши

$$f(z) = \frac{1}{2\pi i} \int_C \frac{f(\zeta)\, d\zeta}{\zeta - z}.$$

Но функция $f(z)$, изображаемая интегралом Коши, дифференцируема в точке z любое число раз, причем

$$f^{(n)}(z) = \frac{n!}{2\pi i} \int_C \frac{f(\zeta)\, d\zeta}{(\zeta - z)^{n+1}},$$

что и доказывает теорему.

Из формулы Коши непосредственно следует такое утверждение:

Если кривая C — окружность, $|\zeta - z| = R$, то, введя замену $\zeta - z = Re^{i\varphi}$, из формулы Коши получаем

$$f(z) = \frac{1}{2\pi} \int_0^{2\pi} f(z + Re^{i\varphi})\, d\varphi.$$

Эта формула выражает теорему о среднем для аналитических функций, которую можно сформулировать так:

Если функция $f(z)$ аналитична внутри круга и непрерывна как в круге, так и на ограничивающей его окружности, то значение функции в центре круга равно среднему арифметическому ее значений на окружности (аналитичность функции на самой окружности здесь не предполагается).

Замечательным свойством аналитической функции является принцип максимума ее модуля.

Модуль функции $f(z)$, аналитической в области D и непрерывной в $\overline{D}$, не может достигать своего наибольшего значения ни в какой точке области D, кроме случая, когда $f(z) \equiv \text{const}$.

Если при указанных условиях еще и $f(z) \neq 0$, то модуль функции $f(z)$ не может достигать в области D и своего наименьшего значения.

Для аналитических функций справедливы следующие теоремы:

Теорема Лиувилля. *Если функция $f(z)$ аналитична во всей плоскости и равномерно ограничена по модулю, то она постоянна.*

Теорема Морера. *Если функция $f(z)$ непрерывна в односвязной области D и интеграл от этой функции, взятый по любому замкнутому контуру C, целиком лежащему в области D, равен нулю $\left(\oint_\sigma f(z)\, dz = 0\right)$, то функция $f(z)$ аналитична в этой области.*

§ 5. РЯДЫ АНАЛИТИЧЕСКИХ ФУНКЦИЙ

Числовые и функциональные ряды. Если $a_k = \alpha_k + \beta_k i$, $k \in N$,— комплексные постоянные, то ряд $\sum\limits_{k=1}^{\infty} a_k = \sum\limits_{k=1}^{\infty} (\alpha_k + \beta_k i)$ есть числовой ряд с комплексными членами (a_k — члены ряда),

$$S_n = \sum_{k=1}^{n} a_k = \sum_{k=1}^{n} (\alpha_k + \beta_k i) = \sum_{k=1}^{n} \alpha_k + i \sum_{k=1}^{n} \beta_k$$

— частная сумма ряда,

$$r_n = \sum_{k=n+1}^{\infty} a_k = \sum_{k=n+1}^{\infty} (\alpha_k + \beta_k i) = \sum_{k=n+1}^{\infty} \alpha_k + i \sum_{k=n+1}^{\infty} \beta_k$$

— остаток ряда.

Если существует предел последовательности частных сумм

$$S = \lim_{n\to\infty} S_n = \lim_{n\to\infty} \sum_{k=1}^{n} \alpha_k + i \lim_{n\to\infty} \sum_{k=1}^{n} \beta_k,$$

то ряд называется с х о д я щ и м с я, а S — суммой ряда.

Очевидно, ряд $\sum\limits_{k=1}^{\infty} a_k$ сходится тогда и только тогда, когда сходится каждый из рядов $\sum\limits_{k=1}^{\infty} \alpha_k$ и $\sum\limits_{k=1}^{\infty} \beta_k$. Этот факт позволяет проводить исследование сходимости рядов с комплексными членами, основываясь на сходимости рядов с действительными членами. Поэтому здесь широко применяются известные из математического анализа функций действительной переменной принцип сравнения рядов, признак д'Аламбера, признак Коши и другие достаточные признаки сходимости рядов.

В случае сходимости ряда $\sum\limits_{k=1}^{\infty} a_k$ его остаток r_n при неограниченном возрастании n стремится к нулю, т. е. для любого $\varepsilon > 0$ можно указать такой номер N, что при $n \geqslant N$ выполняется неравенство $|r_n| < \varepsilon$.

Если ряд $\sum\limits_{k=1}^{\infty} a_k$ сходится, то его общий член a_k при $k \to \infty$ стремится к нулю; $\lim\limits_{k\to\infty} a_k = 0$ (необходимое условие сходимости).

Ряд $\sum\limits_{k=1}^{\infty} a_k$ сходится, если его отрезок $\sum\limits_{k=n}^{n+m} a_k$ при $n \to \infty$ стремится к нулю, т. е. для любого $\varepsilon > 0$ можно указать такой номер N, что $\left|\sum\limits_{k=n}^{n+m} a_k\right| < \varepsilon$, если $n > N$ и $m \geqslant 0$ (достаточное условие сходимости ряда).

Добавление или отбрасывание конечного числа членов ряда не влияет на сходимость ряда.

Если сходится ряд $\sum_{k=1}^{\infty} |a_k|$ с действительными положительными членами, то сходится и ряд $\sum_{k=1}^{\infty} a_k$. При этом последний называется абсолютно сходящимся рядом. В случае абсолютной сходимости ряда $\sum_{k=1}^{\infty} a_k$ $(a_k = \alpha_k + \beta_k i)$ абсолютно сходятся и ряды $\sum_{k=1}^{\infty} \alpha_k$ и $\sum_{k=1}^{\infty} \beta_k$.

Ряд $\sum_{k=1}^{\infty} u_k(z)$, членами которого являются функции комплексной переменной z, называется функциональным рядом. Точка z_0 называется точкой сходимости этого ряда, если сходится числовой ряд $\sum_{k=1}^{\infty} u_k(z_0)$.

Функциональный ряд называется сходящимся в области D, если он сходится в каждой точке этой области. Совокупность всех точек сходимости называется областью сходимости функционального ряда. В общем случае область сходимости может оказаться многосвязной, она может быть и замкнутой.

В области D сходимости функционального ряда $\sum_{k=1}^{\infty} u_k(z)$ можно определить однозначную функцию $f(z)$, значение которой в каждой фиксированной точке z_0 области D равно сумме соответствующего числового ряда: $f(z_0) = \sum_{k=1}^{\infty} u_k(z_0)$. При таком определении функция $f(z)$, называется суммой функционального ряда

$$f(z) = \sum_{k=1}^{\infty} u_k(z),$$

и тогда в любой фиксированной точке $z \in D$ для любого наперед заданного $\varepsilon > 0$ можно указать такой номер N, что $\left| f(z) - \sum_{k=1}^{n} u_k(z) \right| < \varepsilon$ при $n \geqslant N$.

В общем случае номер N зависит от выбора величин ε и z, а если значение N зависит только от ε и одинаково для всех $z \in D$ одновременно, то ряд называется равномерно сходящимся к функции $f(z)$ в области D. Записывают это так:

$$\sum_{k=1}^{\infty} u_k(z) \rightrightarrows f(z).$$

Важен следующий достаточный признак равномерной сходимости функциональных рядов.

Признак Вейерштрасса. *Если в некоторой области D члены ряда, составленного из функций комплексной переменной, не превосходят по модулю соответствующих членов сходящегося числового ряда с положительными членами, то указанный функциональный ряд в области D сходится равномерно, т. е. ряд $\sum_{k=1}^{\infty} u_k(z)$ равномерно сходится в области D, если $|u_k(z)| \leqslant a_k$, $k \in N$, для всех $z \in D$, причем все $a_k > 0$, $a_k = \text{const}$, и ряд $\sum_{k=1}^{\infty} a_k$ сходится.*

Отметим некоторые свойства равномерно сходящихся рядов.

1. *Сумма равномерно сходящегося в области D ряда, состоящего из непрерывных функций, есть функция, непрерывная в области D.*

2. *Равномерно сходящийся в области D ряд непрерывных функций можно почленно интегрировать вдоль всякой кусочно-гладкой кривой C, целиком лежащей в области D; при этом сумма ряда, составленного из интегралов от членов исходного ряда, равна интегралу от суммы исходного ряда вдоль той же кривой, т. е. если в области D все $u_k(z)$, $k \in N$, непрерывны и $\sum_{k=1}^{\infty} u_k(z) \rightrightarrows f(z)$, то*

$$\int_C \left(\sum_{k=1}^{\infty} u_u(z)\right) dz = \sum_{k=1}^{\infty} \int_C u_k(z)\, dz = \int_C f(z)\, dz,$$

где C — любая кусочно-гладкая кривая, целиком лежащая в области D.

3. ***Теорема Вейерштрасса.*** *Ряд, составленный из аналитических в области D функций и равномерно сходящийся в любой замкнутой области $\overline{D}'$, принадлежащей области D, имеет своей суммой аналитическую в области D функцию, его можно почленно дифференцировать произвольное число раз, и ряд, составленный из производных порядка n от членов исходного ряда, равномерно сходится в области $\overline{D}'$ к соответствующей производной от суммы исходного ряда, т. е. если $u_k(z)$, $k \in N$, аналитичны в области D и $\sum_{k=1}^{\infty} u_k(z) \rightrightarrows f(z)$ в области $\overline{D}' \subset D$, то $f(z)$ аналитична в области D, причем, $\frac{d^n}{dz^n}\left(\sum_{k=1}^{\infty} u_k(z)\right) = \sum_{k=1}^{\infty} u_k^{(n)}(z) \rightrightarrows f^{(n)}(z)$ в области $\overline{D}' \subset D$.*

Степенные ряды.

Определение. *Степенным рядом называется ряд вида $\sum_{k=0}^{\infty} c_k (z - z_0)^k$, где c_k — постоянные комплексные числа (коэффициенты ряда).*

Для степенных рядов справедлива следующая теорема.

Теорема Абеля. *Если степенной ряд сходится в точке* z_1, *то он сходится во всех точках* z, *удовлетворяющих условию* $|z - z_0| < |z_1 - z_0|$; *если степенной ряд расходится в точке* z_2, *то он расходится во всех точках* z, *удовлетворяющих условию* $|z - z_0| > |z_2 - z_0|$.

Важны следствия из теоремы Абеля:

Следствия. 1. *Для степенного ряда* $\sum_{k=0}^{\infty} c_k (z - z_0)^k$, *имеющего как точки сходимости (помимо* z_0, *где ряд всегда сходится), так и точки расходимости, всегда существует такое действительное число* $R > 0$, *что внутри круга* $|z - z_0| < R$ *рассматриваемый ряд сходится, а вне этого круга — расходится.*

В круге радиуса $\rho < R$ ряд $\sum_{k=0}^{\infty} c_k (z - z_0)^k$ сходится равномерно.

Область $|z - z_0| < R$ называется кругом сходимости, а число R — радиусом сходимости степенного ряда.

Радиус сходимости R вычисляется по формуле Коши — Адамара $R = \dfrac{1}{\overline{\lim\limits_{k\to\infty}} \sqrt[k]{|c_k|}}$, где $\overline{\lim\limits_{k\to\infty}} \sqrt[k]{|c_k|} = l$ — верхний предел последовательности $\{\sqrt[k]{|c_k|}\}$.

Если $l = \infty$, то считают, что $R = 0$ (ряд сходится лишь в точке z_0). Если $l = 0$, то $R = \infty$ (ряд сходится на всей комплексной плоскости).

2. *Внутри круга сходимости ряд сходится к аналитической функции.*

3. *Степенной ряд внутри круга сходимости (подобно всякому равномерно сходящемуся ряду) можно почленно интегрировать и почленно дифференцировать любое число раз. При этом радиус сходимости каждого вновь полученного ряда равен радиусу сходимости исходного ряда, а над суммой ряда выполняется то же действие, что и над самим рядом.*

4. *Коэффициенты* c_k *степенного ряда могут вычисляться через значения суммы ряда и ее производных, определенных в центре круга сходимости по формулам:* $c_0 = f(z_0)$, $c_k = \dfrac{f^{(k)}(z_0)}{k!}$, $k \in N$.

Пример. Найти сумму ряда $\sum_{k=0}^{\infty} (z - z_0)^k$.

Решение. Каждый коэффициент ряда равен 1, поэтому радиус сходимости $R = 1$. Заданный ряд является рядом геометрической прогрессии, для которого

$$S_n(z) = 1 + (z - z_0) + \cdots + (z - z_0)^n = \frac{1 - (z - z_0)^{n+1}}{1 - (z - z_0)}.$$

Поэтому сумма ряда — аналитическая функция

$$f(z) = \lim_{n\to\infty} S_n(z) = \frac{1}{1-(z-z_0)}.$$

Ряд Тейлора. Выше говорилось о том, что степенной ряд внутри своего круга сходимости определяет аналитическую функцию — сумму этого ряда. Справедливо, однако, и обратное утверждение: *функция $f(z)$, аналитическая внутри круга $|z-z_0|<R$, может быть представлена в этом круге сходящимся степенным рядом* $\sum_{k=0}^{\infty} c_k (z-z_0)^k$, *причем коэффициенты этого ряда однозначно вычисляются по формулам:*

$$c_0 = f(z_0),\ c_k = \frac{f^{(k)}(z_0)}{k!},\quad k \in N.$$

Заметим, что так как $f^{(k)}(z_0) = \frac{k!}{2\pi i}\oint_G \frac{f(\zeta)}{(\zeta-z_0)^{k+1}}d\zeta$, то

$$c_k = \frac{1}{2\pi i}\oint_C \frac{f(\zeta)}{(\zeta-z_0)^{k+1}}d\zeta,$$

где C — некоторый произвольный замкнутый контур внутри круга $|z-z_0|<R$, содержащий точку z_0 внутри.

Разложение аналитической функции в круге $|z-z_0|<R$ в сходящийся степенной ряд $f(z)=\sum_{k=0}^{\infty} c_k (z-z_0)^k$ называется разложением Тейлора, а соответствующий ряд — рядом Тейлора для функции $f(z)$.

Говорят, что функция $f(z)$ голоморфна в точке a, если она в некоторой окрестности этой точки раскладывается в степенной ряд относительно $z-a$. Свойство голоморфности функции в точке a эквивалентно ее аналитичности в этой точке. Так как $f(z)$ голоморфна в точке a, то существует круг с центром в точке a, внутри которого $f(z)$ раскладывается в степенной ряд; $f(z)$ как сумма степенного ряда — аналитическая функция внутри этого круга, следовательно, аналитическая в точке a.

Если же $f(z)$ — аналитическая в точке a, то существует круг с центром в этой точке, внутри которого $f(z)$ будет аналитической функцией. Тогда $f(z)$ может быть представлена в виде суммы степенного ряда, сходящегося внутри этого круга.

Следовательно, $f(z)$ будет голоморфной в точке a.

Функция, голоморфная в каждой точке области D, называется голоморфной в этой области.

Итак, если $f(z)$ голоморфна в области D, то $f(z)$ — аналитична в этой области.

Нули функции. *Если $f(z_0)=0$, то точка z_0 называется нулем этой функции, а разложение функции $f(z)$ в окрестности точки z_0*

в степенной ряд имеет вид

$$f(z) = c_1(z - z_0) + c_2(z - z_0)^2 + \cdots + c_k(z - z_0)^k + \cdots,$$

так как

$$c_0 = f(z_0) = 0.$$

Если не только $c_0 = 0$, но и $c_1 = c_2 = \ldots = c_{k-1} = 0$, а $c_k \neq 0$, то точка z_0 называется нулем k-го порядка функции $f(z)$. В этом случае разложение функции $f(z)$ в окрестности точки z_0 имеет вид

$$f(z) = c_k(z - z_0)^k + c_{k+1}(z - z_0)^{k+1} + \cdots.$$

(Это возможно, когда $f(z_0) = f'(z_0) = \cdots = f^{(k-1)}(z_0) = 0$, а $f^{(k)}(z_0) \neq 0$).

Теорема. *Пусть функция $f(z)$ аналитична в области D и обращается в нуль в различных точках $z_1, z_2, \ldots, z_n$, принадлежащих этой области. Тогда, если существует предел $\lim\limits_{n\to\infty} z_n = a$ и точка a принадлежит области D ($a \in D$), то $f(z) \equiv 0$ в области D.*

Из этой теоремы непосредственно следует, что всякая, не равная тождественно нулю аналитическая в области D функция $f(z)$ имеет в любой замкнутой подобласти $\bar{D}'$ области D ($\bar{D}' \subset D$) лишь конечное число нулей; бесконечное число нулей аналитическая функция может иметь только в открытой (незамкнутой) области или в неограниченной области.

Определение. *Функция $f(z)$, аналитическая во всей комплексной плоскости ($z \neq \infty$), называется целой функцией.*

Из приведенной выше теоремы и ее следствия можно заключить, что целая функция в любой ограниченной области имеет лишь конечное число нулей. Эти нули можно перенумеровать (например, в порядке возрастания их модулей).

На полной комплексной плоскости целая функция может иметь не более чем счетное множество нулей, и предельной точкой этого множества является бесконечно удаленная точка.

Теорема единственности. *Если в области D существует такая последовательность различных точек $z_1, z_2, \ldots, z_n, \ldots$, сходящаяся к точке $a \in D$, что значения двух аналитических в области D функций $f(z)$ и $\varphi(z)$ совпадают во всех точках указанной последовательности, то $f(z) \equiv \varphi(z)$ в области D.* (Нетрудно заметить, что теорема единственности следует из предыдущей теоремы, так как здесь, очевидно, разность $\Phi(z) = f(z) - \varphi(z)$ удовлетворяет условиям последней и потому $\Phi(z) \equiv 0$.)

Теорема единственности играет важную роль в теории функций комплексной переменной. Она, в частности, указывает, что во всякой области D может существовать лишь единственная аналитическая функция, принимающая заданные значения в последовательности точек $z_1, z_2, \ldots, z_n, \ldots$, принадлежащих области D, и сходящейся к точке $a \in D$.

Из этой теоремы следует, что если две аналитические в области D функции $f_1(z)$ и $f_2(z)$ совпадают на некоторой кривой L, принадлежащей области D, то они совпадают и во всей области D.

Из теоремы единственности также можно сделать вывод, что если функция $f_1(z)$ аналитична в области D_1, а функция $f_2(z)$ аналитична в области D_2, причем области D_1 и D_2 имеют общую подобласть D', в которой $f_1(z)$ и $f_2(z)$ совпадают, то существует единственная аналитическая функция $F(z)$ такая, что

$$F(z) = \begin{cases} f_1(z), & z \in D_1, \\ f_2(z), & z \in D_2. \end{cases}$$

В этом случае функция $F(z)$ называется аналитическим продолжением функции $f_1(z)$ (соответственно $f_2(z)$) на область $D_1 + D_2$. При этом функцию $f_2(z)$ также называют аналитическим продолжением функции $f_1(z)$ на область D_2, а функцию $f_1(z)$ — аналитическим продолжением функции $f_2(z)$ на область D_1.

Элементарные функции комплексной переменной. Теорема единственности определения аналитической функции и ее следствия позволяют распространить на комплексную плоскость известные элементарные функции действительной переменной.

Так, если в интервале $[a; b]$ действительной оси Ox задана непрерывная функция $f(x)$, то в области D комплексной плоскости, содержащей отрезок действительной оси $[a; b]$, может существовать единственная аналитическая функция $f(z)$, которая принимает данные значения $f(x)$ на отрезке $[a; b]$. Функция $f(z)$ называется аналитическим продолжением функции $f(x)$ в комплексную область D.

Рассмотрим некоторые часто встречающиеся элементарные функции на действительной оси Ox и их аналитические продолжения в комплексную область.

Для действительной переменной $x \in]-\infty; \infty[$ известны функции:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \quad \sin x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}, \quad \cos x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!},$$

где $0! = 1$.

Нетрудно установить, что аналогичные степенные ряды комплексного аргумента

$$\sum_{k=0}^{\infty} \frac{z^k}{k!}, \quad \sum_{k=0}^{\infty} (-1)^k \frac{z^{2k+1}}{(2k+1)!}, \quad \sum_{k=0}^{\infty} (-1)^k \frac{z^{2k}}{(2k)!}$$

сходятся на всей комплексной плоскости, и, следовательно, их суммы — целые функции z. Эти функции являются аналитическими продолжениями на всю комплексную плоскость функций действительной переменной e^x, $\sin x$ и $\cos x$ соответственно. Для этих продолжений сохранены принятые на действительной оси Ox назва-

ния и обозначения:

$$e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!}, \quad \sin z = \sum_{k=0}^{\infty} (-1)^k \frac{z^{2k+1}}{(2k+1)!}, \quad \cos z = \sum_{k=0}^{\infty} (-1)^k \frac{z^{2k}}{(2k)!}.$$

С помощью функции e^z строятся гиперболические функции комплексного аргумента

$$\operatorname{ch} z = \frac{e^z + e^{-z}}{2} \text{ и } \operatorname{sh} z = \frac{e^z - e^{-z}}{2},$$

которые также являются целыми функциями.

С помощью функций $\sin z$ и $\cos z$ вводятся функции:

$$\operatorname{tg} z = \frac{\sin z}{\cos z} \text{ и } \operatorname{ctg} z = \frac{\cos z}{\sin z}.$$

Известно, что $\ln x = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{(x-1)^k}{k}$, причем этот ряд сходится при всех $x \in]0; 2]$. Поэтому сумма ряда $\sum_{k=1}^{\infty} (-1)^{k-1} \frac{(z-1)^k}{k}$ представляет собой аналитическое продолжение функции $\ln x$ в область $|z-1| < 1$; ей даны то же название «логарифм» и то же обозначение $\ln z$. Функция $\ln z$ аналитична в круге $|z-1| < 1$.

Подобно этому строится аналитическая в круге $|z| < 1$ функция

$$\arcsin z = \sum_{k=0}^{\infty} \frac{1 \cdot 3 \cdot 5 \ldots (2k-1)}{2^k k! (2k+1)} z^{2k+1},$$

являющаяся аналитическим продолжением функции $\arcsin x$ действительного аргумента при $x \in [-1; 1]$.

Заметим, что функции $\ln z$ и $\arcsin z$ не являются целыми функциями, так как определяющие их ряды сходятся не на всей комплексной плоскости, а лишь внутри соответствующих кругов единичного радиуса.

Теорема единственности определения аналитических функций распространяется и на соотношения между функциями. Это позволяет строить также аналитические продолжения в комплексную плоскость соотношений между функциями действительной переменной.

Так, в комплексной области будут справедливы соотношения:

$$\sin^2 z + \cos^2 z = 1, \quad \cos(z_1 + z_2) = \cos z_1 \cos z_2 - \sin z_1 \sin z_2,$$

$$\sin 2z = 2 \sin z \cos z, \quad e^{\ln z} = z, \quad e^{z_1} \cdot e^{z_2} = e^{z_1 + z_2} \text{ и т. д.}$$

Более того, оказывается, что аналитические продолжения в комплексную плоскость справедливы и для дифференциальных соотношений между функциями. Это приводит к тому, что если некоторая функция $f(x)$ действительной переменной представляет собой решение какого-то дифференциального уравнения и имеет аналитическое

продолжение в область D комплексной плоскости, то это ее аналитическое продолжение будет решением дифференциального уравнения, которое является аналитическим продолжением в ту же область D указанного дифференциального уравнения.

Остановимся теперь на отдельных свойствах некоторых из рассмотренных выше элементарных функций комплексной переменной z.

Прежде всего установим связь между показательной и тригонометрическими функциями z.

В данном параграфе функция e^z определена с помощью ряда и в то же время, как было показано для $z = x + iy$ (см. с. 18), она определена иначе: $e^z = e^x(\cos y + i \sin y)$.

Для доказательства равносильности этих определений в формуле

$$e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!}$$

положим $z = i\zeta$, тогда

$$e^{i\zeta} = \sum_{k=0}^{\infty} i^k \frac{\zeta^k}{k!} = \sum_{l=0}^{\infty} (-1)^l \frac{\zeta^{2l}}{(2l)!} + i \sum_{l=0}^{\infty} (-1)^l \frac{\zeta^{2l+1}}{(2l+1)!}.$$

Отсюда получим, что для всех значений комплексной переменной $z = i\zeta$

$$e^{i\zeta} = \cos\zeta + i \sin\zeta.$$

Эта зависимость называется формулой (тождеством) Эйлера (§ 1, гл. I).

При $z = x + iy$ имеем $e^z = e^{x+iy} = e^x \cdot e^{iy}$, так как $e^{z_1+z_2} = e^{z_1} \cdot e^{z_2}$, и, следовательно, $e^z = e^x(\cos y + i \sin y)$.

Из формулы Эйлера непосредственно следует, что $e^{-i\zeta} = \cos\zeta - i \sin\zeta$.

Используя выражения для $e^{i\zeta}$ и $e^{-i\zeta}$, находим, что

$$\cos z = \frac{1}{2}(e^{iz} + e^{-iz}), \qquad \sin z = \frac{1}{2i}(e^{iz} - e^{-iz}).$$

Если в этих формулах положить iz вместо z, то получим:

$$\cos iz = \operatorname{ch} z, \quad \sin iz = i \operatorname{sh} z.$$

Справедливы и обратные соотношения:

$$\operatorname{ch} iz = \cos z, \quad \operatorname{sh} iz = i \sin z.$$

Для функции $w = e^z = e^x(\cos y + i \sin y)$ имеем:

$$|w| = e^x, \quad \arg w = y.$$

Так как $\cos y = \cos(y + 2k\pi)$, $\sin y = \sin(y + 2k\pi)$, то при $z = x + iy$

$$e^z = e^x(\cos y + i \sin y) = e^x(\cos(y + 2k\pi) + i \sin(y + 2k\pi)) =$$
$$= e^x e^{i(y+2k\pi)} = e^{(x+iy)+i2k\pi} = e^{z+2k\pi i},$$

т. е. функция e^z периодическая с периодом $i2k\pi$, поэтому $y = \arg w$ является лишь главным значением аргумента функции $w = e^z$, а полный аргумент

$$\operatorname{Arg} w = \arg w + 2k\pi.$$

Теперь рассмотрим функцию $\int_1^z \frac{d\zeta}{\zeta} = \ln z$, являющуюся аналитическим продолжением $\ln x$ на комплексную плоскость с разрезом вдоль отрицательной части действительной оси Ox.

Известно, что для действительных $x > 0$ имеет место равенство $e^{\ln x} = x$. Поэтому в области $-\pi < \arg z < \pi$ справедливо соотношение $e^{\ln z} = z$.

Следовательно, функция $\ln z$ является обратной для функции e^z.

Функция $\ln z$ — это лишь главное значение логарифма. Так как $e^z = e^{z+i2\pi k}$, то логарифм является функцией многозначной. Полное выражение логарифма z обозначают $\operatorname{Ln} z$, поэтому

$$\operatorname{Ln} z = \ln z + i2k\pi.$$

Пусть $\ln z = u + iv$, тогда $z = e^{\ln z} = e^{u+iv} = e^u e^{iv}$; $|z| = e^u$, $u = \ln|z|$, $\arg z = v$, а потому $\ln z = \ln|z| + i\arg z$;

$$\operatorname{Ln} z = \ln|z| + i\operatorname{Arg} z = \ln|z| + i\arg z + i2k\pi.$$

В частности,

$$\ln i = \ln|i| + i\arg i = \ln 1 + i\frac{\pi}{2} = i\frac{\pi}{2},$$

$$\operatorname{Ln} i = i\frac{\pi}{2} + i2k\pi.$$

Элементарные функции комплексной переменной встречаются в теории электромагнитного поля. В частности, с помощью функций $\operatorname{sh} z$, $\operatorname{ch} z$, $\operatorname{th} z$ описываются сложные явления поверхностного эффекта в плоском электропроводящем листе и эффекта близости в двух близко расположенных друг от друга параллельных шинах (проводах), по которым протекают синусоидальные во времени токи. Поверхностный эффект в плоском проводящем листе возникает при прохождении переменного магнитного потока сквозь поперечное сечение листа и выражается в неравномерном распределении интенсивности магнитного поля по сечению. В результате магнитное поле как бы оттесняется от середины сечения листа к его боковой поверхности. Эффект близости проявляется в неравномерном распределении плотности электрического тока по поперечным сечениям проводов.

Ряд Лорана. *Ряд вида* $\sum_{k=-\infty}^{\infty} c_k (z - z_0)^k$, *где* k *принимает все целые (положительные, отрицательные и нулевое) значения, называется ряд о м Л о р а н а.*

Этот ряд понимается как сумма двух рядов:

$$\sum_{k=-\infty}^{\infty} c_k (z - z_0)^k = \sum_{k=0}^{\infty} c_k (z - z_0)^k + \sum_{k=1}^{\infty} c_{-k} (z - z_0)^{-k}.$$

Первая часть ряда Лорана — ряд, расположенный по неотрицательным степеням двучлена $(z - z_0)$ и рассмотренный выше, называется правильной частью, а вторая часть — ряд, расположенный по отрицательным степеням двучлена $(z - z_0)$, — главной частью ряда Лорана.

Областью сходимости ряда Лорана является общая часть области сходимости его главной части и области сходимости его правильной части. Известно, что область сходимости правильной части ряда Лорана — это круг радиуса $R_1 = 1 : \overline{\lim} \sqrt[k]{|c_k|}$ с центром в точке z_0. Внутри этого круга ряд сходится к некоторой аналитической функции $f_1(z)$ — сумме ряда. Для определения области сходимости главной части ряда Лорана введем обозначение $(z - z_0)^{-1} = \xi$, тогда получим ряд $\sum_{k=1}^{\infty} c_{-k} \xi^k$. Это степенной ряд, сходящийся в круге радиуса $\rho = \frac{1}{\overline{\lim}_{k \to \infty} \sqrt[k]{|c_{-k}|}}$ к аналитической функции, которую назовем $\varphi(\xi)$:

$$\varphi(\xi) = \sum_{k=1}^{\infty} c_{-k} \xi^k, \quad |\xi| < \rho.$$

Вернемся к переменной z:

$$\varphi\left(\frac{1}{z - z_0}\right) = \sum_{k=1}^{\infty} c_{-k} (z - z_0)^{-k}, \quad \left|\frac{1}{z - z_0}\right| < \rho \text{ или } |z - z_0| > \frac{1}{\rho}.$$

Обозначим: $\varphi\left(\frac{1}{z - z_0}\right) = f_2(z)$ и $\frac{1}{\rho} = R_2$. Получим

$$f_2(z) = \sum_{k=1}^{\infty} c_{-k} (z - z_0)^{-k}, \quad |z - z_0| > R_2,$$

т. е. ряд, сходящийся к аналитической функции $f_2(z)$ вне круга радиуса R_2 с центром в точке z_0.

Итак,

$$f_1(z) = \sum_{k=0}^{\infty} c_k (z - z_0)^k, \quad |z - z_0| < R_1;$$

$$f_2(z) = \sum_{k=1}^{\infty} c_{-k} (z - z_0)^{-k}, \quad |z - z_0| > R_2.$$

Если $R_1 > R_2$ (рис. 13), то существует общая область сходимости рядов, составляющих ряд Лорана. Это круговое кольцо $R_2 < |z - z_0| < R_1$. Внутри этого кольца ряд $\sum_{k=-\infty}^{\infty} c_k (z - z_0)^k$ схо-

дится к некоторой аналитической функции

$$f(z) = f_1(z) + f_2(z).$$

При $R_2 > R_1$ ряд Лорана нигде не сходится.

Теорема. *Функция $f(z)$, аналитическая в круговом кольце $R_2 < |z - z_0| < R_1$, однозначно представляется в этом кольце рядом Лорана.*

Коэффициенты ряда Лорана вычисляются по формулам, подобным формулам для нахождения коэффициентов ряда Тейлора:

$$c_n = \frac{1}{2\pi i} \oint\limits_{C_{k'}} \frac{f(\xi)}{(\xi - z_0)^{n+1}}\, d\xi, \quad n \in Z,$$

где $C_{k'}$ — любой замкнутый контур в кольце $R_2 < |z - z_0| < R_1$, содержащий точку z_0 внутри.

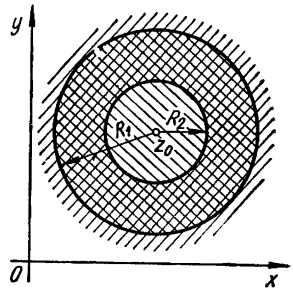

Рис. 13

Если функция $f(z)$ однозначна и аналитична в окрестности бесконечно удаленной точки (эту окрестность можно рассматривать как круговое кольцо с центром в начале координат и ограниченное окружностью некоторого конечного радиуса R и окружностью бесконечного радиуса), то в этой окрестности $f(z)$ может быть разложена в ряд Лорана.

Действительно, если ввести новую комплексную переменную ξ по формуле $z = \frac{1}{\xi}$, то $f(z) = f\left(\frac{1}{\xi}\right) = \varphi(\xi)$ и функция $\varphi(\xi)$ будет аналитической функцией в окрестности точки $\xi = 0$, причем

$$\lim_{z \to \infty} f(z) = \lim_{\xi \to 0} \varphi(\xi).$$

Теперь ряд Лорана для функции $f(z)$ в окрестности бесконечно удаленной точки можно получить из лорановского разложения функции $\varphi(\xi)$ в окрестности точки $\xi = 0$ заменой переменной $\xi = \frac{1}{z}$. Это приведет к тому, что правильная часть ряда Лорана для $\varphi(\xi)$ будет главной для $f(z)$, а главная часть ряда для $\varphi(\xi)$ будет правильной для $f(z)$.

Правильные и особые точки аналитической функции. При изучении функций действительной переменной мы встречались с различным поведением функции в той или иной точке: с непрерывностью, конечным разрывом, устранимым разрывом, бесконечным разрывом. Подобные точки встречаются и для аналитических функций комплексной переменной. Рассмотрим такие точки.

Точка z_0 называется правильной точкой функции $f(z)$, если существует степенной ряд $\sum\limits_{k=0}^{\infty} c_k (z - z_0)^k$, который сходится к функции $f(z)$ в круге сколь угодно малого радиуса с центром

в точке z_0, т. е. если существует круг сколь угодно малого радиуса с центром в точке z_0, в котором функция $f(z)$ аналитична.

Точки, не являющиеся правильными, называются особыми точками функции $f(z)$.

Если функция $f(z)$ аналитична в области D, то все внутренние точки этой области — правильные точки функции $f(z)$. Среди точек границы Γ области D могут быть как правильные, так и особые точки функции $f(z)$. Если все точки области D и ее границы Γ — правильные точки, то функция $f(z)$ аналитична в замкнутой области $\bar{D}$. В этом случае функцию $f(z)$ можно аналитически продолжить на некоторую другую область D', содержащую область $\bar{D}$, а потому функцию $f(z)$ можно продолжить аналитически и через любой участок границы, не содержащий особых точек. Аналитическое продолжение через участок границы, целиком состоящий из особых точек, невозможно.

Теорема. *Если степенной ряд в некотором круге сходится к функции $f(z)$, то на границе круга сходимости есть хотя бы одна особая точка этой аналитической функции $f(z)$.*

Из этой теоремы следует, что радиус круга сходимости степенного ряда $\sum_{k=0}^{\infty} c_k (z - z_0)^k$ равен расстоянию точки z_0 до ближайшей особой точки суммы этого ряда.

Точка z_0 называется изолированной особой точкой функции $f(z)$, если $f(z)$ — функция однозначная и аналитическая в кольце $0 < |z - z_0| < R_1$, а точка z_0 — особая точка функции $f(z)$. В самой точке z_0 функция $f(z)$ может быть и не определена.

Очевидно, функцию $f(z)$ можно в кольце $0 < |z - z_0| < R_1$ разложить в ряд Лорана. При этом возможны три случая:

1) Ряд Лорана не содержит членов с отрицательными степенями двучлена $(z - z_0)$, т. е.

$$f(z) = \sum_{k=0}^{\infty} c_k (z - z_0)^k = c_0 + c_1 (z - z_0) + c_2 (z - z_0)^2 + \cdots$$

Здесь $\lim\limits_{z \to z_0} f(z) = c_0$. Доопределим функцию $f(z)$, положив $f(z_0) = c_0$, тогда $\lim\limits_{z \to z_0} f(z) = c_0 = f(z_0)$. При таком доопределении функция $f(z)$ уже не будет иметь разрыва в точке z_0.

В этом случае точка z_0 называется устранимой особой точкой функции $f(z)$.

2) Ряд Лорана содержит конечное число m членов с отрицательными степенями двучлена $(z - z_0)$:

$$f(z) = \sum_{k=-m}^{\infty} c_k (z - z_0)^k = c_{-m} (z - z_0)^{-m} + c_{-(m-1)} (z - z_0)^{-(m-1)} +$$

$$+ \cdots + c_{-1} (z - z_0)^{-1} + c_0 + c_1 (z - z_0) + c_2 (z - z_0)^2 + \cdots$$

В этом случае точка z_0 называется полюсом порядка m функции $f(z)$.

Если z_0 — полюс функции $f(z)$, то при $z \to z_0$ всегда $|f(z)|$ неограниченно возрастает, независимо от способа стремления z к z_0. Поэтому здесь функцию $f(z)$ можно записать в виде

$$f(z) = \frac{\psi(z)}{(z - z_0)^m},$$

где $\psi(z)$ — аналитическая функция, причем $\psi(z_0) \neq 0$.

Обозначим: $\frac{(z - z_0)^m}{\psi(z)} = g(z)$. Тогда, очевидно, если точка z является нулем порядка m функции $g(z)$, то для функции $f(z) = \frac{1}{g(z)}$ она будет полюсом порядка m и, наоборот, если точка z_0 — полюс порядка m для функции $f(z)$, то она будет нулем порядка m для функции $g(z)$.

3) Ряд Лорана содержит бесконечное число членов с отрицательными степенями двучлена $(z - z_0)$.

В таком случае точка z_0 называется существенно особой точкой функции $f(z)$. В такой точке аналитическая функция $f(z)$ не имеет ни конечного, ни бесконечного предела (ее предел не существует в этой точке). Однако, выбрав определенную последовательность точек $z_n \to z_0$, можно получить соответствующую последовательность значений функции, сходящуюся к произвольно заданному комплексному числу. Меняя последовательности точек вблизи z_0, можно получать различные последовательности соответствующих значений функции, сходящиеся к различным пределам. Это и есть отличительное свойство существенно особой точки.

Если бесконечно удаленная точка не является правильной точкой однозначной аналитической функции $f(z)$, то она будет изолированной особой точкой этой функции тогда, когда существует окрестность точки $z = \infty$, в которой нет особых точек функции $f(z)$, кроме самой бесконечно удаленной точки.

Бесконечно удаленная изолированная особая точка так же, как и изолированная особая точка в конечной области, может оказаться одной из трех видов — устранимой особой точкой, полюсом порядка m или существенно особой точкой.

В заключение отметим, что теоремы о поведении однозначной аналитической функции вблизи ее изолированных особых точек носят обратимый характер.

Кроме определения целой функции, приведенного выше при рассмотрении нулей функции, укажем второе определение: *функция $f(z)$ называется целой (или голоморфной), если она не имеет особых точек в конечной области.*

Если функция $f(z)$ не имеет других особых точек, кроме полюсов, то она называется дробной (или мероморфной). Мероморфными функциями являются, например, все целые функции, дробно-рациональные и тригонометрические.

§ 6. ВЫЧЕТЫ

Если z_0 — изолированная особая точка однозначной аналитической в области D функции $f(z)$, то в окрестности этой точки функцию $f(z)$ можно представить рядом Лорана:

$$f(z) = \sum_{n=-\infty}^{\infty} c_n (z - z_0)^n,$$

где

$$c_n = \frac{1}{2\pi i} \int_{\gamma} \frac{f(\xi)\, d\xi}{(\xi - z_0)^{n+1}}, \quad n \in \mathbf{Z},$$

γ — любой замкнутый контур, лежащий в области D аналитичности функции $f(z)$ и охватывающий точку z_0, но не содержащий внутри других особых точек этой функции.

Вычетом аналитической функции $f(z)$ в изолированной особой точке z_0 называется коэффициент c_{-1} указанного ряда Лорана, стоящий при $(z - z_0)^{-1}$, т. е. значение интеграла

$$\frac{1}{2\pi i} \int_{\gamma} f(\xi)\, d\xi,$$

вычисленного в положительном направлении по замкнутому контуру γ.

Обозначается вычет функции $f(z)$ в изолированной особой точке z_0 одним из двух символов: Выч $[f(z); z_0]$ или Res $[f(z); z_0]$. Далее мы будем пользоваться первым из этих обозначений.

Непосредственно из определения следует, что в правильной точке и в устранимой особой точке вычет функции равен нулю.

Если особая точка — полюс, то вычисление вычета просто. Возможны два случая:

1) z_0 — полюс первого порядка аналитической функции. Тогда ряд Лорана для функции $f(z)$ в окрестности точки z_0 имеет вид

$$f(z) = \frac{c_{-1}}{z - z_0} + c_0 + c_1 (z - z_0) + c_2 (z - z_0)^2 + \dots,$$

откуда

$$(z - z_0) f(z) = c_{-1} + c_0 (z - z_0) + c_1 (z - z_0)^2 + \dots$$

и потому

$$\text{Выч}\,[f(z); z_0] = c_{-1} = \lim_{z \to z_0} ((z - z_0) f(z)).$$

Этому результату можно придать другую форму.

Действительно, в случае полюса первого порядка в точке z_0 функция $f(z)$ в окрестности этой точки может быть представлена как отношение двух функций $\varphi(z)$ и $\psi(z)$:

$$f(z) = \frac{\varphi(z)}{\psi(z)},$$

где $\varphi(z_0) \neq 0$, а $\psi(z)$ имеет в точке z_0 нуль первого порядка: $\psi(z_0) = 0$, $\psi'(z_0) \neq 0$.

Так как для функции $\psi(z)$ ряд Тейлора в окрестности точки z_0 имеет вид

$$\psi(z) = \psi(z_0) + \frac{\psi'(z_0)}{1!}(z - z_0) + \frac{\psi''(z_0)}{2!}(z - z_0)^2 + \frac{\psi'''(z_0)}{3!}(z - z_0)^3 + \ldots,$$

то

$$\psi(z) = (z - z_0)\left(\psi'(z_0) + \frac{\psi''(z_0)}{2!}(z - z_0) + \frac{\psi'''(z_0)}{3!}(z - z_0)^2 + \ldots\right).$$

Поэтому Выч $[f(z);\ z_0] = \lim\limits_{z \to z_0}((z - z_0) f(z)) = \dfrac{\varphi(z_0)}{\psi'(z_0)}$.

Пример. Найти вычеты функции $f(z) = \dfrac{z^2+1}{z^2-z-2}$ в ее особых точках.

Р е ш е н и е. Особые точки заданной функции $f(z)$ — это полюсы первого порядка $z_0 = -2$ и $z_1 = 1$; $f(z)$ можно представить в виде $f(z) = \dfrac{\varphi(z)}{\psi(z)}$, где $\varphi(z) = z^2 + 1$, $\psi(z) = z^2 - z - 2 = (z - 1)(z + 2)$; $\psi'(z) = 2z - 1$.

Поэтому

$$\text{Выч}\,[f(z);\ -2] = \frac{\varphi(-2)}{\psi'(-2)} = \frac{(-2)^2 + 1}{2 \cdot (-2) - 1} = -1,$$

$$\text{Выч}\,[f(z);\ 1] = \frac{\varphi(1)}{\psi'(1)} = 2.$$

2) z_0 — полюс порядка m функции $f(z)$.

В этом случае ряд Лорана для функции $f(z)$ в окрестности точки z_0 имеет вид

$$f(z) = \sum_{n=-m}^{\infty} c_n (z - z_0)^n = \frac{c_{-m}}{(z - z_0)^m} + \frac{c_{-m+1}}{(z - z_0)^{m-1}} + \cdots$$

$$\cdots + \frac{c_{-2}}{(z - z_0)^2} + \frac{c_{-1}}{z - z_0} + c_0 + c_1 (z - z_0) + c_2 (z - z_0)^2 + \ldots,$$

откуда

$$(z - z_0)^m f(z) = c_{-m} + c_{-m+1}(z - z_0) + c_{-m+2}(z - z_0)^2 + \cdots$$

$$\cdots + c_{-1}(z - z_0)^{m-1} + c_0 (z - z_0)^m + c_1 (z - z_0)^{m+1} + \ldots.$$

Дифференцируя обе части этого равенства $m - 1$ раз по переменной z, получим

$$\frac{d^{m-1}}{dz^{m-1}}((z - z_0)^m f(z)) = (m - 1)!\, c_{-1} + m(m - 1) \ldots 2c_0 (z - z_0) +$$

$$+ (m + 1)\, m (m - 1) \ldots 3c_1 (z - z_0)^2 +$$

$$+ (m + 2)(m + 1)\, m (m - 1) \ldots 4c_2 (z - z_0)^3 + \ldots,$$

поэтому

$$\text{Выч}\,[f(z);\ z_0] = \frac{1}{(m-1)!}\lim_{z\to z_0}\frac{d^{m-1}}{dz^{m-1}}\left((z-z_0)^m f(z)\right).$$

Пример. Найти вычеты функции $f(z) = \dfrac{e^z}{(z-1)^2(z+2)^3}$ в ее особых точках.

Решение. Особыми точками функции $f(z)$ являются: $z_0 = -2$ — полюс третьего порядка, $z_1 = 1$ — полюс второго порядка.

Следовательно,

$$\text{Выч}\,[f(z);\ -2] = \frac{1}{2!}\lim_{z\to -2}\frac{d^2}{dz^2}\left(\frac{e^z}{(z-1)^2}\right) =$$

$$= \frac{1}{2!}\lim_{z\to -2}\frac{(z^2-6z+11)\,e^z}{(z-1)^4} = \frac{1}{6e^2},$$

$$\text{Выч}\,[f(z);\ 1] = \lim_{z\to 1}\frac{d}{dz}\left(\frac{e^z}{(z+2)^3}\right) = \lim_{z\to 1}\frac{(z-1)\,e^z}{(z+2)^4} = 0.$$

Основная теорема о вычетах. *Если функция $f(z)$ однозначна и аналитична в замкнутой области D всюду, за исключением конечного числа изолированных особых точек z_k $(k = 1, 2, 3, \ldots, n)$, лежащих внутри области D, то*

$$\int_{\Gamma+} f(\xi)\,d\xi = 2\pi i\sum_{k=1}^{n}\text{Выч}\,[f(z);\ z_k],$$

где Γ^+ — полная граница области D, проходимая в положительном направлении.

Доказательство. Окружим каждую изолированную особую точку z_k функции $f(z)$ достаточно малым замкнутым контуром γ_k $(k = 1, 2, 3, \ldots, n)$, целиком лежащим в области D и не содержащим других особых точек функции $f(z)$ (рис. 14).

Внутри образовавшейся многосвязной области, ограниченной контуром Γ и всеми контурами γ_k $(k = 1, 2, 3, \ldots, n)$, функция $f(z)$ аналитична. Поэтому по теореме Коши

$$\int_{\Gamma+} f(\xi)\,d\xi + \sum_{k=1}^{n}\int_{\gamma_k^-} f(\xi)\,d\xi = 0.$$

Отсюда

$$\int_{\Gamma+} f(\xi)\,d\xi = -\sum_{k=1}^{n}\int_{\gamma_k^-} f(\xi)\,d\xi = \sum_{k=1}^{n}\int_{\gamma_k^+} f(\xi)\,d\xi = 2\pi i\sum_{k=1}^{n}\text{Выч}\,[f(z);\ z_k].$$

Определение. *Вычетом аналитической функции $f(z)$ в бесконечно удаленной изолированной особой точке называется коэффициент ряда Лорана в окрестности этой*

точки при z^{-1}, взятый с противоположным знаком, т. е.

$$\text{Выч } [f(z);\ \infty] = -c_{-1} = -\int\limits_{\gamma+} f(\xi)\,d\xi = +\int\limits_{\gamma-} f(\xi)\,d\xi,$$

где γ — окружность с центром в начале координат и достаточно большого радиуса R (такого, что при $|z| > R$ нет других особых точек функции $f(z)$, кроме бесконечно удаленной точки), а интегрирование ведется при отрицательном обходе, так как только при таком обходе (по ходу часовой стрелки), окрестность бесконечно удаленной точки всегда остается слева.

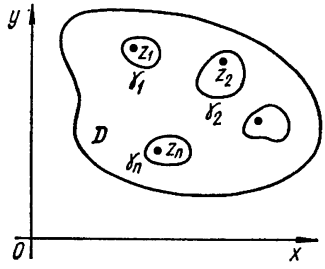

Рис. 14

Приведенное определение и правило вычисления вычета в бесконечно удаленной изолированной особой точке позволяют сформулировать следующую теорему.

Теорема. *Если функция $f(z)$ аналитична на полной комплексной плоскости всюду, кроме конечного числа n изолированных особых точек, среди которых есть бесконечно удаленная точка, то сумма всех вычетов этой функции в изолированных особых точках равна нулю.*

Доказательство. Пусть C — замкнутый контур, внутри которого находится $n-1$ изолированная особая точка $z_1, z_2, \ldots, z_{n-2}, z_{n-1}$ функции $f(z)$. Тогда по основной теореме о вычетах

$$\frac{1}{2\pi i}\int\limits_{C+} f(\xi)\,d\xi = \sum_{k=1}^{n-1} \text{Выч } [f(z);\ z_k],$$

но

$$\frac{-1}{2\pi i}\int\limits_{C+} f(\xi)\,d\xi = -\text{Выч } [f(z);\ \infty] = -\text{Выч } [f(z);\ z_n]\ (z_n = \infty).$$

Следовательно, $\sum\limits_{k=1}^{n} \text{Выч } [f(z);\ z_k] = 0$. Теорема доказана.

Эта теорема часто используется для упрощения вычисления интеграла от функции комплексной переменной по замкнутому контуру.

Примеры

1. Вычислить $\int\limits_0^{2\pi} \frac{d\varphi}{1 + a\cos\varphi}$ $(|a| < 1)$.

Решение. Пусть $e^{i\varphi} = z$, тогда

$$dz = ie^{i\varphi}d\varphi = iz d\varphi,\quad d\varphi = \frac{dz}{iz}.$$

При изменении φ от нуля до 2π комплексная переменная $z = e^{i\varphi}$ пробегает окружность $|z| = 1$ в положительном направлении. Так как по формуле Эйлера

$$\cos\varphi = \frac{1}{2}(e^{i\varphi} + e^{-i\varphi}) = \frac{1}{2}\left(z + \frac{1}{z}\right),$$

то

$$\int_0^{2\pi} \frac{d\varphi}{1 + a\cos\varphi} = \frac{1}{i}\int_{|z|=1} \frac{1}{1 + \frac{a}{2}\left(z + \frac{1}{z}\right)} \cdot \frac{dz}{z} = \frac{2}{i}\int_{|z|=1} \frac{dz}{az^2 + 2z + a}.$$

Особыми точками подынтегральной функции $f(z) = \dfrac{1}{az^2 + 2z + a}$ являются полюсы первого порядка, они совпадают с нулями знаменателя

$$z_{1,2} = \frac{-1 \pm \sqrt{1 - a^2}}{a}.$$

Так как здесь $z_1 z_2 = 1$, то лишь полюс $z_1 = \dfrac{-1 + \sqrt{1 - a^2}}{a}$ лежит внутри круга $|z| < 1$, поэтому

$$\int_0^{2\pi} \frac{d\varphi}{1 + a\cos\varphi} = 2\pi i \frac{2}{i} \operatorname{Выч}[f(z); z_1] = 4\pi \lim_{z \to z_1} (f(z)(z - z_1)) =$$

$$= 4\pi \lim_{z \to z_1} \frac{z - z_1}{a(z - z_1)(z - z_2)} = \frac{4\pi}{a(z - z_2)}\Big|_{z=z_1} = \frac{4\pi}{a(z_1 - z_2)} = \frac{2\pi}{\sqrt{1 - a^2}}.$$

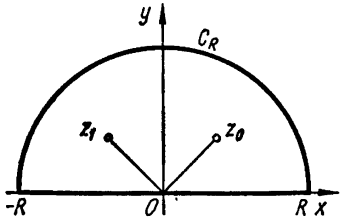

Рис. 15

2. Вычислить $\displaystyle\int_{-\infty}^{+\infty} \frac{dx}{x^4 + 1}$.

Решение. Подынтегральную функцию $\dfrac{1}{x^4 + 1}$ действительной переменной будем рассматривать, как аналитическую функцию $\dfrac{1}{z^4 + 1}$ комплексной переменной, заданную на действительной оси: $\dfrac{1}{x^4 + 1} = \dfrac{1}{z^4 + 1}\Big|_{z=x}$. Для решения поставленной задачи используем аналитическое продолжение этой функции на верхнюю полуплоскость и вычислим интеграл от комплексной переменной $f(z) = \dfrac{1}{z^4 + 1}$ по некоторому замкнутому контуру, содержащему ось абсцисс ($-\infty < x < +\infty$). В качестве такого контура выберем контур C, состоящий из отрезка $[-R; R]$ действительной оси и полуокружности C_R при $R > 1$ (рис. 15).

Знаменатель функции $f(z)$ имеет четыре нуля $z_k = e^{\frac{i(\pi + 2k\pi)}{4}}$ ($k = 0, 1, 2, 3$) — это полюсы функции $f(z)$, два из которых ($z_0 = e^{i\frac{\pi}{4}}$ и $z_1 = e^{i\frac{3\pi}{4}}$) расположены в верхней полуплоскости. Поэтому выбранный нами контур интегрирования C охватывает только два полюса z_0 и z_1 функции $f(z)$.

Следовательно,

$$\int_C \frac{dz}{z^4 + 1} = \int_C f(\zeta)\, d\zeta = 2\pi i \sum_{k=0}^{1} \operatorname{Выч}[f(z); z_k],$$

при этом $f(z)=\dfrac{\varphi(z)}{\psi(z)}$, где $\varphi(z)\equiv 1$, $\psi(z)=z^4+1$, $\psi'(z)=4z^3$. Так как Выч $[f(z);\ z_k]=\dfrac{\varphi(z_k)}{\psi'(z_k)}=\dfrac{1}{4z_k^3}$ $(k=0,\ 1)$, то

$$\int_C f(\zeta)\,d\zeta=2\pi i\left(\frac{1}{4z_0^3}+\frac{1}{4z_1^3}\right)=\frac{\pi i}{2}\left(e^{-i\frac{3\pi}{4}}+e^{-i\frac{9\pi}{4}}\right)=$$

$$=\frac{\pi i}{2}\left(\left(\cos\frac{3\pi}{4}-i\sin\frac{3\pi}{4}\right)+\left(\cos\frac{9\pi}{4}-i\sin\frac{9\pi}{4}\right)\right)=$$

$$=\frac{\pi}{2}\left(i\left(\cos\frac{3\pi}{4}+\cos\frac{\pi}{4}\right)+\left(\sin\frac{3\pi}{4}+\sin\frac{\pi}{4}\right)\right)=\frac{\pi}{2}\cdot\frac{2}{\sqrt{2}}=\frac{\pi}{\sqrt{2}}.$$

Таким образом, $\int\limits_C f(\zeta)\,d\zeta=\int\limits_{-R}^{R}\dfrac{dx}{x^4+1}+\int\limits_{C_R} f(\zeta)\,d\zeta=\dfrac{\pi}{\sqrt{2}}$.

Перейдем в этом равенстве к пределу при $R\to\infty$. При достаточно большом $R>R_0>1$ получим: $|f(z)|<\dfrac{M}{R^4}$, где R_0 и M — положительные постоянные. Поэтому $\left|\int\limits_{C_R} f(\zeta)\,d\zeta\right|<\dfrac{M}{R^4}\int\limits_{C_R} d\zeta=\dfrac{M\pi R}{R^4}=\dfrac{\pi M}{R^3}$ и, следовательно, $\lim\limits_{R\to\infty}\int\limits_{C_R} f(\zeta)\,d\zeta=0$.

Окончательно получаем: $\lim\limits_{R\to\infty}\left(\int\limits_{-R}^{R}\dfrac{dx}{x^4+1}+\int\limits_{C_R} f(\zeta)\,d\zeta\right)=\int\limits_{-\infty}^{\infty}\dfrac{dx}{x^4+1}$, откуда

$$\int_{-\infty}^{\infty}\frac{dx}{x^4+1}=\frac{\pi}{\sqrt{2}}.$$

Логарифмический вычет. Пусть однозначная функция $f(z)$ аналитична всюду в области D, за исключением конечного числа p изолированных особых точек z_k $(k=1, 2, 3, \ldots, p)$, причем все особые точки z_k являются полюсами. Пусть, кроме того, функция $f(z)$ в области D имеет конечное число n нулей; z_k $(k=1, 2, 3, \ldots, n)$, а на границе Γ области D нет ни нулей, ни особых точек функции $f(z)$.

Полным числом N нулей (полным числом P полюсов) функции $f(z)$, расположенных в области D, называется количество всех ее нулей (полюсов) в этой области при условии, что каждый нуль (полюс) считается столько раз, каков его порядок.

Функция $\varphi(z)=\dfrac{f'(z)}{f(z)}=(\ln f(z))'_z$ называется логарифмической производной функции $f(z)$, а вычеты функции

$\varphi(z)$ в ее особых точках z_m ($m = 1, 2, 3, \ldots, M$) — логарифмическими вычетами функции $f(z)$.

Особыми точками функции $\varphi(z)$ являются нули и полюсы функции $f(z)$. Если точка $\tilde{z}_k$ является нулем порядка (кратности) n_k функции $f(z)$, то в окрестности этой точки

$$f(z) = (z - \tilde{z}_k)^{n_k} f_1(z), \quad \text{где } f_1(z) \neq 0.$$

Поэтому

$$\varphi(z) = \frac{f'(z)}{f(z)} = (\ln f(z))'_z = (n_k \ln(z - \tilde{z}_k) + \ln f_1(z))'_z =$$

$$= \frac{n_k}{z - \tilde{z}_k} + \frac{f_1'(z)}{f_1(z)}$$

и точка $\tilde{z}_k$ будет полюсом первого порядка для функции $\varphi(z)$.

Следовательно, $\text{Выч}\,[\varphi(z);\ \tilde{z}_k] = \text{Выч}\left[\frac{f'(z)}{f(z)};\ \tilde{z}_k\right] = n_k$, т. е. логарифмический вычет функции $f(z)$ относительно ее нуля равен порядку (кратности) этого нуля.

Если z_k — полюс порядка (кратности) p_k функции $f(z)$, то для функции $F(z) = \frac{1}{f(z)}$ точка z_k будет нулем того же порядка p_k, и так как $\ln f(z) = -\ln F(z)$, то

$$\frac{f'(z)}{f(z)} = (\ln f(z))'_z = (-\ln F(z))'_z = -(\ln F(z))' = -\frac{F'(z)}{F(z)}.$$

Это означает, что $\text{Выч}\left[\frac{f'(z)}{f(z)};\ z_k\right] = -\text{Выч}\left[\frac{F'(z)}{F(z)};\ z_k\right] = -p_k$, т. е. логарифмический вычет функции $f(z)$ относительно ее полюса равен порядку (кратности) этого полюса, взятому со знаком минус.

Полученные результаты позволяют записать формулы для определения полного числа N нулей и полного числа P полюсов функции $f(z)$ в виде:

$$N = \sum_{k=1}^{n} \text{Выч}\left[\frac{f'(z)}{f(z)};\ \tilde{z}_k\right], \quad P = \sum_{k=1}^{p} \text{Выч}\left[\frac{f'(z)}{f(z)};\ z_k\right].$$

В соответствии с основной теоремой о вычетах теперь получаем:

$$\frac{1}{2\pi i} \int_{\Gamma+} \frac{f'(\xi)}{f(\xi)}\, d\xi = \sum_{k=1}^{n} \text{Выч}\left[\frac{f'(z)}{f(z)};\ \tilde{z}_k\right] + \sum_{k=1}^{p} \text{Выч}\left[\frac{f'(z)}{f(z)};\ z_k\right],$$

т. е. $\frac{1}{2\pi i} \int_{\Gamma+} \frac{f'(\xi)}{f(\xi)}\, d\xi = N - P$ (выражение $\frac{1}{2\pi i} \int_{\Gamma} \frac{f'(z)}{f(z)}\, dz$ назы-

вают логарифмическим вычетом функции $f(z)$ относительно контура Γ).

Теорема. *Если $f(z)$ — однозначная функция, аналитическая в области D всюду, кроме конечного числа изолированных особых точек, причем все эти точки являются полюсами, а на границе Γ области D эта функция не имеет ни нулей, ни особых точек, то логарифмический вычет функции $f(z)$ относительно замкнутого контура Γ равен разности между полным числом нулей и полным числом полюсов, расположенных внутри этого контура.*

В соответствии с определением функции комплексной переменной $w = w(z)$ всегда справедливо следующее утверждение:

Если независимая комплексная переменная z получает некоторое приращение Δz, а точка z комплексной плоскости Z обходит какой-то контур Γ, то и функция $w = w(z)$ получает некоторое приращение $\Delta w = w(z + \Delta z) - w(z)$, а точка w комплексной плоскости W обходит какой-то соответствующий контур G. При этом аргумент комплексной функции $w = w(z)$ также получает некоторое свое приращение.

Этот факт позволяет найти логарифмический вычет функции (число $N - P$).

Действительно, если для функции $f(z)$ в области D, ограниченной замкнутым контуром Γ, выполнены все условия, наложенные теоремой о логарифмическом вычете, то

$$\frac{1}{2\pi i}\int\limits_{\Gamma+}\frac{f'(\xi)}{f(\xi)}d\xi = \frac{1}{2\pi i}\int\limits_{\Gamma+} d\ln f(\xi) = \frac{1}{2\pi i}\int\limits_{\Gamma+} d(\ln|f(\xi)| +$$

$$+ i\arg f(\xi)) = \frac{1}{2\pi i}\int\limits_{\Gamma+} d\ln|f(\xi)| + \frac{1}{2\pi}\int\limits_{\Gamma+} d\arg f(\xi).$$

Так как $\ln|f(\xi)|$ — однозначная функция, то ее приращение при полном обходе точкой ξ замкнутого контура Γ равно нулю, т. е. $\int\limits_{\Gamma+} d\ln|f(\xi)| = 0$.

Поэтому

$$\frac{1}{2\pi i}\int\limits_{\Gamma+}\frac{f'(\xi)}{f(\xi)}d\xi = N - P = \frac{1}{2\pi}\int\limits_{\Gamma+} d\arg f(\xi) = \frac{1}{2\pi}\operatorname{Var}[\arg f(z)]_{\Gamma+},$$

где $\operatorname{Var}[\arg f(z)]_{\Gamma+}$ — полное приращение аргумента комплексной функции $w = f(z)$ при обходе точкой z замкнутого контура Γ в положительном направлении.

Таким образом, число $N - P = \frac{1}{2\pi}\operatorname{Var}[\arg f(z)]_{\Gamma+}$ равно числу оборотов, совершаемых точкой w вокруг точки $w = 0$. Этот вывод часто позволяет упростить нахождение числа нулей аналитической функции внутри заданной области.

Существенную роль при этом играет следующая теорема.

Теорема Руше. *Если на границе Γ области D имеет место неравенство $|f(z)|_\Gamma > |\varphi(z)|_\Gamma$, причем функции $f(z)$ и $\varphi(z)$ аналитичны в замкнутой области $\bar{D}$, то в области D полное число нулей функции $F(z) = f(z) + \varphi(z)$ равно полному числу нулей функции $f(z)$.*

Пример. Применяя теорему Руше, найти полное число нулей функции $w = z^2 - 1$ внутри круга $|z| \leqslant 2$.

Р е ш е н и е. Если положить $z^2 = f(z)$, $-1 = \varphi(z)$, то на границе Γ окружности $|z| = 2$ получим

$$|f(z)|_\Gamma = 4,\ |\varphi(z)|_\Gamma = 1,\ |f(z)|_\Gamma > |\varphi(z)|_\Gamma.$$

По теореме Руше внутри круга $|z| \leqslant 2$ функция $w = F(z) = f(z) + \varphi(z) = z^2 - 1$ имеет столько же нулей, сколько их имеет функция $f(z) = z^2$, но эта функция имеет один нуль $z = 0$ второй кратности, т. е. ее полное число нулей равно двум.

Следовательно, функция $w = z^2 - 1$ внутри круга $|z| \leqslant 2$ имеет два нуля.

С помощью теоремы Руше можно сравнительно просто доказать основную теорему алгебры, утверждающую, что всякая целая функция (многочлен) имеет по крайней мере один нуль.

В электротехнике вычеты функции используются при синтезе электрических цепей, когда решают задачи нахождения структуры схемы и определения параметров входящих в нее элементов по заданным частотным или временным свойствам этой схемы цепи.

§ 7. ОСНОВНЫЕ СВЕДЕНИЯ О КОНФОРМНОМ ОТОБРАЖЕНИИ

Выясним геометрический смысл аргумента и модуля производной аналитической функции.

Пусть $w = f(z)$ — аналитическая функция в некоторой области D на комплексной плоскости Z.

Проведем через любую точку $z_0 \in D$ произвольную гладкую кривую γ_1, лежащую в D.

Функция $w = f(z)$ преобразует точку z_0 в точку $w_0 = f(z_0)$ на комплексной плоскости W. При этом кривая γ_1 преобразуется в некоторую кривую Γ_1, лежащую в плоскости W и проходящую через точку w_0. Предположим, что $f'(z_0) \neq 0$. Тогда

$$f'(z_0) = \lim_{\substack{\Delta z \to 0 \\ (z \to z_0)}} \frac{\Delta w}{\Delta z} = Ke^{i\alpha} = K(\cos\alpha + i\sin\alpha),$$

где

$$\alpha = \arg f'(z_0), \quad K = |f'(z_0)| \neq 0.$$

Так как при делении комплексных чисел аргументы вычитаются, то, на основании определения α, имеем:

$$\alpha = \lim_{\Delta z \to 0} \arg \frac{\Delta w}{\Delta z} = \lim_{\Delta z \to 0} (\arg \Delta w - \arg \Delta z) =$$

$$= \lim_{\Delta z \to 0} \arg \Delta w - \lim_{\Delta z \to 0} \arg \Delta z = \Phi_1 - \varphi_1,$$

где Φ_1 и φ_1 — углы, которые образуют касательные к кривым Γ_1 и γ_1 соответственно в точках w_0 и z_0 с положительными направлениями действительных полуосей на комплексных плоскостях W и Z.

Для другой пары кривых Γ_2 и γ_2 в тех же точках w_0 и z_0 имеем

$$\Phi_2 - \varphi_2 = \alpha.$$

Следовательно, $\alpha = \Phi_1 - \varphi_1 = \Phi_2 - \varphi_2$, и этот угол зависит только от функции $w = f(z)$ и точки z_0 (соответственно w_0).

Теперь очевидно, что $\Phi_1 - \Phi_2 = \varphi_1 - \varphi_2 = \theta$, т. е. если кривые γ_1 и γ_2 образуют в точке z_0 на плоскости Z угол θ, то такой же угол θ будут образовывать в точке w_0 кривые Γ_1 и Γ_2 — отображения кривых γ_1 и γ_2 на плоскости W (рис. 16). Для модуля производной, очевидно,

$$K = \lim_{\substack{\Delta z \to 0 \\ (z \to z_0)}} \left| \frac{\Delta w}{\Delta z} \right| = \lim_{\substack{\Delta z \to 0 \\ (w \to w_0)}} \frac{|\Delta w|}{|\Delta z|},$$

т. е. K — это предел отношения длин векторов Δw и Δz при перемещении точек w и z соответственно по кривым Γ и γ в точки w_0 и z_0. Этот предел зависит только от функции $f(z)$ и точки z_0 (соответственно w_0).

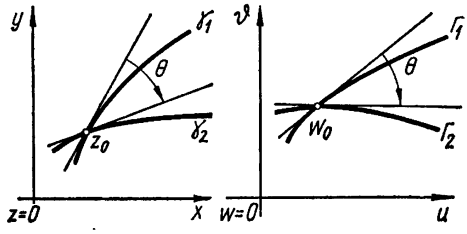

Рис. 16

Итак, аналитическая функция $f(z)$ осуществляет отображение области D плоскости Z на область G плоскости W таким образом, что углы между кривыми в точке z_0 на плоскости Z и между соответствующими кривыми на плоскости W одинаковы и, кроме того, достаточно малые линейные элементы в точке z_0 плоскости Z преобразуются подобно (пропорционально) в достаточно малые элементы в точке w_0 плоскости W, причем $|f'(z_0)|$ определяет коэффициент преобразования подобия. Эти свойства отображения называются соответственно свойством сохранения углов и свойством постоянства растяжения.

Определение. *Отображение $w = f(z)$, сохраняющее углы между кривыми и постоянство растяжений, называется конформным отображением (конформным преобразованием).*

Если при этом сохраняется и направление отсчета углов, то такое отображение называется конформным отображением первого рода; если направление отсчета углов изменяется на противоположное,— конформным отображением второго рода.

Заметим, что всякое отображение, осуществляемое аналитической функцией $w = f(z)$, конформно во всех точках, где $f'(z) \neq 0$.

В то же время, отмечая, что при конформном отображении точки плоскости Z переходят в точки плоскости W, линии переходят в линии, замкнутые линии — в замкнутые линии, надо обратить внимание на то, что область на плоскости Z не всегда преобразуется в

область на плоскости W. Конформное отображение области D комплексной плоскости Z осуществляется на область G комплексной плоскости W только в том случае, когда аналитическая функция $w = f(z)$ однолистна в области D, т. е. в различных точках этой области она принимает различные значения. В этом случае отображение является взаимно однозначным. Если функция $w = f(z)$ не однолистна, то отображение не будет взаимно однозначным, а следовательно, не может быть конформным (напомним, что аналитическая функция всегда однозначна).

Например, функция $w = z^4$ в области D — полукольце $1 \leqslant |z| \leqslant 2$, $0 \leqslant \arg z \leqslant \pi$, аналитична, и в этой области $w' = 4z^3 \neq 0$. Однако заданная функция отображает область D на об-

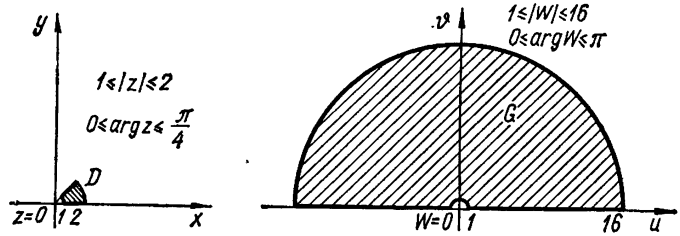

Рис. 17

ласть $1 \leqslant |w| \leqslant 16$, $0 \leqslant \arg w \leqslant 4\pi$, представляющую собой дважды покрытое кольцо $1 \leqslant |w| \leqslant 16$ на плоскости W, функция $w = z^4$ не однолистна; таким образом преобразование, осуществляемое функцией $w = z^4$ в этом примере, не является конформным.

В то же время та же функция $w = z^4$ отображает область $1 \leqslant |z| \leqslant 2$, $0 \leqslant \arg z \leqslant \frac{\pi}{4}$ на область $1 \leqslant |w| \leqslant 16$, $0 \leqslant \arg w \leqslant \pi$ конформно (рис. 17).

Выше предполагалось, что производится конформное отображение ограниченной области D на ограниченную область G. Однако встречаются случаи, когда приходится рассматривать отображение окрестности точки $z = \infty$ на окрестность точки $w = \infty$ (или наоборот). Это отображение считается конформным, если окрестность точки z_0 конформно отображается на окрестность точки $\xi = 0$, где $\xi = \frac{1}{w}$. Подобно этому определяется конформное отображение окрестности точки $z = \infty$ на окрестность точки $w = \infty$.

В заключение рассмотрим несколько примеров конформных преобразований.

Примеры

1. Преобразование с помощью линейной функции $w = az + b$, где a и b — постоянные комплексные числа, причем $a \neq 0$.

Эта функция существует при всех z и однозначна. Обратная функция $z = \frac{w - b}{a}$ существует при всех w и также однозначна. Следовательно, линейная

функция $w = az + b$ однолистна во всех точках плоскости Z, она устанавливает взаимно однозначное соответствие между точками плоскости Z и точками плоскости W.

Обозначим $\arg a = \alpha$, $\arg z = \varphi$ и рассмотрим сперва случай, когда $b = 0$.

При $b = 0$ $w = az = |a| \cdot |z| \cdot e^{i(\alpha+\varphi)}$. Отсюда делаем вывод, что это преобразование состоит в подобном растяжении в $|\alpha|$ раз и в повороте плоскости вокруг точки $z = 0$ на угол $\alpha = \arg a$.

При $b \neq 0$ к этому преобразованию добавится еще сдвиг точек плоскости на $|b|$ в направлении вектора $\boldsymbol{b}$, изображающего комплексное число b.

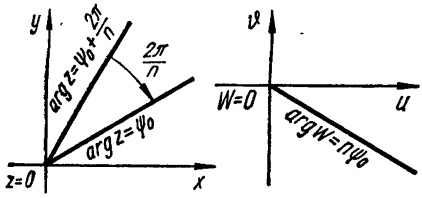

Рис. 18

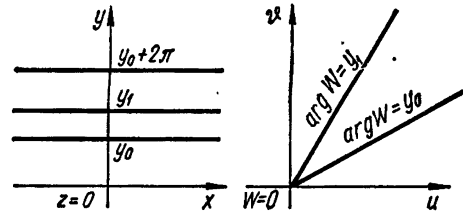

Рис. 19

Таким образом, линейная функция $w = az + b$ преобразует комплексную плоскость Z в комплексную плоскость W путем подобного растяжения, поворота и сдвига; и, так как функция $w = az + b$ однолистна, а $w' = a \neq 0$, то это преобразование конформно.

2. Преобразование с помощью степенной функции $w = z^n$, $n > 1$ — целое.

Функция $w = z^n$ однолистна в секторе $\psi_0 < \arg z < \psi_0 + \frac{2\pi}{n}$, ее производная $w' = nz^{n-1}$ во всех точках, кроме $z = \infty$, существует и при $z \neq 0$ отлична от нуля. Следовательно, при $z \neq 0$ и $z \neq \infty$ отображение конформно.

Так как при $z = |z| \cdot e^{i\varphi}$ будет $z^n = |z|^n \cdot e^{in\varphi}$, то с помощью функции $w = z^n$ весь указанный сектор перейдет в $n\psi_0 < \arg w < n\psi_0 + 2\pi$, т. е. во всю плоскость, разрезанную по лучу $\arg w = n\psi_0$ (рис. 18). Если сектор меньше указанного, то, естественно, он будет отражен лишь на часть плоскости. Так, в частности, функция $w = z^2$ отображает первую координатную четверть $0 < \arg z < \frac{\pi}{2}$ комплексной плоскости Z на верхнюю полуплоскость W ($\operatorname{Im} w > 0$). Эта же функция $w = z^2$ отображает всю верхнюю полуплоскость Z ($\operatorname{Im} z > 0$) на плоскость W, разрезанную по действительной положительной полуоси. В точках $z = 0$ и $z = \infty$ конформность отображения нарушается.

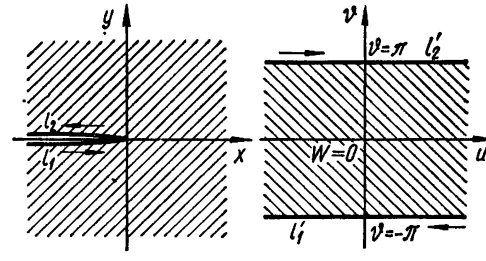

Рис. 20

3. Преобразование с помощью функции $w = e^z$. Нетрудно проверить, что в случае $y_0 \leqslant \operatorname{Im} z < y_0 + 2\pi$ эта функция однолистна, а так как $w' = e^z \neq 0$ и существует при всех z, то отображение, осуществляемое этой функцией в указанной полосе, конформно.

При этом, если $y_0 < y_1 < y_0 + 2\pi$, то полупрямая $y = y_1 = \text{const}$ ($x < 0$) преобразуется функцией $w = e^z$ в единичный отрезок $[OA$ ($|OA| = 1$) на луче $\arg w = y_1$ (рис. 19); полупрямая $y = y_1 = \text{const}$ ($x \geqslant 0$) — в остальную часть луча $\arg w = y_1$. Следовательно, полоса $y_0 < \operatorname{Im} z < y_0 + 2\pi$ на плоскости Z преобразуется в плоскость W, разрезанную по лучу $\arg w = y_0$.

4. Построить функцию, отображающую плоскость Z с разрезом вдоль отрицательной части вещественной оси ($-\infty < \operatorname{Re} z < 0$, $\operatorname{Im} z = 0$) на полосу $-\pi < v < \pi$ в плоскости W (при $w = u + iv$).

Р е ш е н и е. В предыдущем примере было установлено, что функция $w = e^z$ конформно отображает полосу $y_0 < \operatorname{Im} z < y + 2\pi$ в плоскость с разрезом по лучу $\arg w = y_0$. По-видимому, здесь надо рассмотреть обратное отображение — отображение полосы $v_0 \leqslant v \leqslant v_0 + 2\pi$ плоскости W при условии $v_0 = -\pi$ на всю плоскость Z с разрезом по лучу $\arg z_0 = v_0 = -\pi$. Такое отображение реализует функция $z = e^w$. Поэтому необходимое по условию задачи отображение дает функция $w = \ln z = \ln |z| + i \arg z$.

Проверим правильность решения задачи на границе области (рис. 20). Когда точка z пробегает по нижнему краю l_1 разреза от $x = -\infty$ до $x = 0$, то в плоскости W соответствующая точка описывает линию l_1' ($v = -\pi$) от $u = +\infty$ до $u = -\infty$, т. е. всю нижнюю границу полосы $-\pi \leqslant v \leqslant \pi$. Когда точка z пробегает по верхнему краю l_2 разреза от $x = 0$ до $x = -\infty$, то в плоскости W соответствующая точка описывает линию l_2' ($v = -\pi + 2\pi = \pi$) от $u = -\infty$ до $u = +\infty$, т. е. всю верхнюю границу той же полосы. Поэтому область D в плоскости Z и соответствующая ей область G в плоскости W остаются при обходе контуров с одной и той же стороны (справа),— иначе говоря, направление отсчета углов сохраняется. Здесь имеет место конформное отображение первого рода.

Методы конформных отображений широко используются при исследовании плоских электростатических и магнитостатических полей.

ГЛАВА II

ИНТЕГРАЛ ФУРЬЕ

В электротехнике, в автоматике и особенно в радиотехнике и в технике связи довольно часто встречаются электрические напряжения, токи, магнитные потоки и другие величины, изменяющиеся во времени по несинусоидальному непериодическому закону. В этих случаях для исследования процессов, протекающих в электро- и радиотехнических устройствах, прибегают к разложению таких функций в интеграл Фурье.

В данной главе рассматриваются лишь те сведения по интегралу Фурье, которые необходимы инженеру электро- и радиотехнической специальности на практике.

§ 8. ИНТЕГРАЛ ФУРЬЕ

Известно, что если $f(x)$ — непрерывная $2l$-периодическая функция, удовлетворяющая условиям, при которых ее можно разложить в ряд Фурье с периодом $2l$, то

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{l} + b_n \sin \frac{n\pi x}{l} \right), \qquad (1)$$

где

$$a_0 = \frac{1}{l}\int_{-l}^{l} f(t)\,dt,$$

$$a_n = \frac{1}{l}\int_{-l}^{l} f(t)\cos\frac{n\pi t}{l}\,dt,$$

$$b_n = \frac{1}{l}\int_{-l}^{l} f(t)\sin\frac{n\pi t}{l}\,dt.$$

Здесь сумма ряда (1) есть $2l$-периодическая функция, совпадающая на всей числовой оси с данной функцией $f(x)$.

Если $f(x)$ не является $2l$-периодической функцией, но удовлетворяет условиям, когда ее можно разложить в ряд Фурье с периодом $2l$, то сумма ряда Фурье вне $[-l; l]$ не будет совпадать с $f(x)$.

Для того чтобы и в этом случае получить представление $f(x)$ в виде (1), справедливое для всех x, предположим, что в любом конечном промежутке $[-l; l]$ выполняются условия, при которых $f(x)$ можно разложить в ряд Фурье с периодом $2l$, и что $\int_{-\infty}^{+\infty} |f(x)|\,dx = I$ сходится (т. е. $f(x)$ абсолютно интегрируема на всей числовой оси).

Найдем предел выражения, стоящего в правой части равенства (1), при $l \to +\infty$.

Подставив значения a_0, a_n и b_n в правую часть равенства (1), получим

$$f(x) = \frac{1}{2l}\int_{-l}^{l} f(t)\,dt + \frac{1}{l}\sum_{n=1}^{\infty}\int_{-l}^{l} f(t)\left(\cos\frac{n\pi x}{l}\cos\frac{n\pi t}{l} + \right.$$

$$\left. + \sin\frac{n\pi x}{l}\sin\frac{n\pi t}{l}\right)dt = \frac{1}{2l}\int_{-l}^{l} f(t)\,dt +$$

$$+ \frac{1}{l}\sum_{n=1}^{\infty}\int_{-l} f(t)\cos\frac{n\pi}{l}(x-t)\,dt. \qquad (2)$$

Покажем, что первое слагаемое в правой части этого равенства стремится к нулю при $l \to +\infty$.

$$\left|\frac{1}{2l}\int_{-l}^{l} f(t)\,dt\right| = \frac{1}{2l}\left|\int_{-l}^{l} f(t)\,dt\right| \leqslant \frac{1}{2l}\int_{-l}^{l} |f(t)|\,dt,$$

но

$$\int_{-l}^{l} |f(t)|\,dt \leqslant \int_{-\infty}^{+\infty} |f(t)|\,dt = I,$$

тогда

$$\left|\frac{1}{2l}\int_{-l}^{l} f(t)\,dt\right| \leqslant \frac{l}{2l}.$$

Следовательно, $\frac{1}{2l}\int_{-l}^{l} f(t)\,dt \to 0$ при $l \to +\infty$.

Рассмотрим второе слагаемое в равенстве (2).

Пусть переменная величина ω принимает значения

$$\omega_1 = \frac{\pi}{l},\quad \omega_2 = \frac{2\pi}{l},\quad \ldots,\quad \omega_n = \frac{n\pi}{l},\quad \ldots,$$

при этом

$$\Delta\omega_n = \omega_n - \omega_{n-1} = \frac{\pi}{l} \to 0 \text{ при } l \to +\infty,$$

$$\frac{1}{l}\sum_{n=1}^{\infty}\int_{-l}^{l} f(t)\cos\frac{n\pi}{l}(x-t)\,dt = \frac{1}{\pi}\sum_{n=1}^{\infty}\left(\int_{-l}^{l} f(t)\cos\frac{n\pi}{l}(x-t)\,dt\right)\frac{\pi}{l}\,dt = \frac{1}{\pi}\sum_{n=1}^{\infty}\left(\int_{-l}^{l} f(t)\cos\omega_n(x-t)\,dt\right)\Delta\omega_n.$$

Полученное выражение по своей структуре является интегральной суммой для функции

$$\Phi(\omega) = \frac{1}{\pi}\int_{-l}^{l} f(t)\cos\omega(x-t)\,dt.$$

Поэтому, переходя в равенстве (2) к пределу при $l \to +\infty$, получим

$$f(x) = \frac{1}{\pi}\lim_{l\to+\infty}\sum_{n=1}^{\infty}\left(\int_{-l}^{l} f(t)\cos\omega_n(x-t)\,dt\right)\Delta\omega_n,$$

или

$$f(x) = \frac{1}{\pi}\int_{0}^{+\infty} d\omega\int_{-\infty}^{+\infty} f(t)\cos\omega(x-t)\,dt. \qquad (3)$$

Формула (3) называется формулой Фурье, а двойной несобственный интеграл

$$\frac{1}{\pi}\int_{0}^{+\infty} d\omega\int_{-\infty}^{+\infty} f(t)\cos\omega(x-t)\,dt$$

— интегралом Фурье для функции $f(x)$.

Замечание. Рассуждения, в результате которых получена формула Фурье, не являются достаточно строгими. Однако строго доказывается, что если функция $f(x)$ удовлетворяет на каждом конечном интервале условиям Дирихле и абсолютно интегрируема на бесконечном интервале, то формула (3) справедлива. Предполагается, что $f(x)$ в точке x непрерывна.

Формулу (3) можно записать так:

$$f(x) = \int_0^{+\infty} \left(\left(\frac{1}{\pi} \int_{-\infty}^{+\infty} f(t) \cos \omega t dt \right) \cos \omega x + \right.$$

$$\left. + \left(\frac{1}{\pi} \int_{-\infty}^{+\infty} f(t) \sin \omega t dt \right) \sin \omega x \right) d\omega.$$

Введем обозначения:

$$A(\omega) = \frac{1}{\pi} \int_{-\infty}^{+\infty} f(t) \cos \omega t dt,$$

$$B(\omega) = \frac{1}{\pi} \int_{-\infty}^{+\infty} f(t) \sin \omega t dt,$$

тогда

$$f(x) = \int_0^{+\infty} (A(\omega) \cos \omega x + B(\omega) \sin \omega x) d\omega \qquad (4)$$

или

$$f(x) = \int_0^{+\infty} K(\omega) \sin (\omega x - \delta(\omega)) d\omega, \qquad (5)$$

где

$$K(\omega) = \sqrt{A^2(\omega) + B^2(\omega)},$$

$$\delta(\omega) = \operatorname{arc\,tg} \frac{B(\omega)}{A(\omega)}.$$

Из формулы (5) следует, что функция $f(x)$ представляется интегралом Фурье в виде «суммы» гармоник (с амплитудами $K(\omega)$ и начальными фазами $\delta(\omega)$, круговые частоты которых ω изменяются непрерывно в интервале $[0; +\infty[$).

В связи с тем что при разложении функции $f(x)$ в ряд Фурье с периодом $2l$ круговые частоты гармоник имеют дискретные значения $\omega_1 = \frac{\pi}{l}$, $\omega_2 = \frac{2\pi}{l}$, ..., $\omega_n = \frac{n\pi}{l}$, ..., говорят о точечном спектре частот ряда Фурье и о непрерывном спектре частот интеграла Фурье.

Если для периодической функции амплитудный и фазовый спектры являются линейчатыми, то для непериодической функции, представленной интегралом Фурье, они получаются сплошными.

Известно, что в ряд Фурье можно раскладывать и функции, имеющие в некоторых точках разрывы первого рода. Сумма ряда Фурье в указанных точках равна $\frac{1}{2}(f(x-0) + f(x+0))$, где $f(x-0)$ и $f(x+0)$ — левый и правый пределы функции $f(x)$.

При разложении функции $f(x)$ в интеграл Фурье также допускается наличие точек разрыва первого рода, поэтому в левых частях формул (4) и (5) вместо $f(x)$ в общем случае пишут

$$\frac{f(x-0)+f(x+0)}{2}.$$

Если $f(x)$ в точке x непрерывна, то, очевидно,

$$\frac{f(x-0)+f(x+0)}{2}=f(x).$$

§ 9. НЕКОТОРЫЕ ЧАСТНЫЕ СЛУЧАИ ПРЕДСТАВЛЕНИЯ ФУНКЦИИ ИНТЕГРАЛОМ ФУРЬЕ

1. Пусть $f(x)$ — четная функция.

Тогда $f(x)\cdot\cos\omega x$ также четная функция, а $f(x)\cdot\sin\omega x$ — нечетная функция x.

В этом случае

$$A(\omega)=\frac{1}{\pi}\int_{-\infty}^{+\infty}f(t)\cos\omega t dt=\frac{2}{\pi}\int_{0}^{+\infty}f(t)\cos\omega t dt,$$

$$B(\omega)=\frac{1}{\pi}\int_{-\infty}^{+\infty}f(t)\sin\omega t dt=0.$$

Формула Фурье (4) для четной функции $f(x)$ примет вид

$$\frac{f(x-0)+f(x+0)}{2}=\frac{2}{\pi}\int_{0}^{+\infty}\cos\omega x d\omega\int_{0}^{+\infty}f(t)\cos\omega t dt. \qquad (6)$$

2. Пусть $f(x)$ — нечетная функция.

Тогда $f(x)\cdot\cos\omega x$ — нечетная, а $f(x)\cdot\sin\omega x$ — четная функция x. В этом случае

$$A(\omega)=0,$$

$$B(\omega)=\frac{2}{\pi}\int_{0}^{+\infty}f(t)\sin\omega t dt.$$

Формула Фурье (4) для нечетной функции $f(x)$ запишется так:

$$\frac{f(x-0)+f(x+0)}{2}=\frac{2}{\pi}\int_{0}^{+\infty}\sin\omega x d\omega\int_{0}^{+\infty}f(t)\sin\omega t dt. \qquad (7)$$

3. Пусть $f(x)$ задана лишь в $[0;+\infty[$.

Доопределим ее в $]-\infty;0[$ с помощью функции $\varphi(x)$.

Получим функцию

$$F(x)=\begin{cases}\varphi(x) & \text{при } x<0,\\ f(x) & \text{при } x\geqslant 0,\end{cases}$$

заданную в $]-\infty;+\infty[$.

Интеграл Фурье для функции $F(x)$ будет иметь вид

$$\frac{F(x-0)+F(x+0)}{2}=\frac{1}{\pi}\int_0^{+\infty}d\omega\int_{-\infty}^{+\infty}F(t)\cos\omega(x-t)\,dt=$$

$$=\frac{1}{\pi}\int_0^{+\infty}d\omega\int_{-\infty}^{0}\varphi(t)\cos\omega(x-t)\,dt+\frac{1}{\pi}\int_0^{+\infty}d\omega\int_0^{+\infty}f(t)\cos\omega(x-t)\,dt, \tag{8}$$

где

$$\frac{F(x-0)+F(x+0)}{2}=\begin{cases}\dfrac{\varphi(x-0)+\varphi(x+0)}{2} & \text{при } x<0,\\ \dfrac{\varphi(0-0)+f(0+0)}{2} & \text{при } x=0,\\ \dfrac{f(x-0)+f(x+0)}{2} & \text{при } x>0.\end{cases}$$

Если $\varphi(x)$ выбрать так, чтобы $F(x)$ стала четной или нечетной, то формула (8) примет вид (6) или (7) соответственно.

§ 10. КОМПЛЕКСНАЯ ФОРМА ИНТЕГРАЛА ФУРЬЕ. ПРЕОБРАЗОВАНИЕ ФУРЬЕ

Пусть $f(x)$ абсолютно интегрируема на всей числовой оси. Так как $|f(t)\sin\omega(x-t)|\leqslant|f(t)|$, то $\int_{-\infty}^{+\infty}f(t)\sin\omega(x-t)\,dt$ сходится (абсолютно и равномерно) и представляет собой непрерывную функцию ω. Очевидно, при замене ω на $-\omega$ эта функция меняет лишь знак, т. е. является нечетной функцией ω. Тогда

$$\int_{-\infty}^{+\infty}d\omega\int_{-\infty}^{+\infty}f(t)\sin\omega(x-t)\,dt=0.$$

Аналогично получим, что

$$\int_{-\infty}^{+\infty}f(t)\cos\omega(x-t)\,dt$$

является непрерывной и четной функцией ω. Следовательно,

$$\int_{-\infty}^{+\infty}d\omega\int_{-\infty}^{+\infty}f(t)\cos\omega(x-t)\,dt=2\int_0^{+\infty}d\omega\int_{-\infty}^{+\infty}f(t)\cos\omega(x-t)\,dt,$$

при этом

$$\frac{1}{\pi}\int_0^{+\infty}d\omega\int_{-\infty}^{+\infty}f(t)\cos\omega(x-t)\,dt=\frac{1}{2\pi}\int_{-\infty}^{+\infty}d\omega\int_{-\infty}^{+\infty}f(t)\cos\omega(x-t)\,dt-$$

$$-\frac{1}{2\pi} i \int_{-\infty}^{+\infty} d\omega \int_{-\infty}^{+\infty} f(t) \sin \omega (x-t)\, dt =$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega \int_{-\infty}^{+\infty} f(t) (\cos \omega (x-t) - i \sin \omega (x-t))\, dt =$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega \int_{-\infty}^{+\infty} f(t) \cdot e^{-i\omega(x-t)} dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega \int_{-\infty}^{+\infty} f(t) \cdot e^{i\omega(t-x)} dt.$$

В результате формулу Фурье (3)

$$\frac{f(x-0)+f(x+0)}{2} = \frac{1}{\pi} \int_{0}^{+\infty} d\omega \int_{-\infty}^{+\infty} f(t) \cos \omega (x-t)\, dt$$

можно записать так:

$$\frac{f(x-0)+f(x+0)}{2} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega \int_{-\infty}^{+\infty} f(t) \cdot e^{i\omega(t-x)} dt. \qquad (9)$$

Это комплексная форма интеграла Фурье. Предполагая здесь $f(x)$ непрерывной, получим

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-i\omega x} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t)\, e^{i\omega t} dt \right) d\omega. \qquad (10)$$

Пусть

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t) \cdot e^{i\omega t} dt. \qquad (11)$$

Эта функция называется преобразованием Фурье для функции $f(x)$. Теперь формула (10) примет вид

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} F(\omega) \cdot e^{-i\omega x} d\omega. \qquad (12)$$

Эту функцию называют обратным преобразованием Фурье для функции $F(\omega)$.

Равенство (12) при заданной функции $f(x)$ можно рассматривать как интегральное уравнение с неизвестной функцией $F(\omega)$ (так как $F(\omega)$ находится под знаком интеграла). Тогда формула (11) дает решение этого интегрального уравнения.

Аналогично равенство (11) можно рассматривать как интегральное уравнение относительно функции $f(x)$. Тогда его решение дается формулой (12).

Полученные выше формулы (6) и (7) для непрерывной функции $f(x)$, когда она четная или нечетная, можно соответственно запи-

:ать так:

$$f(x) = \sqrt{\frac{2}{\pi}} \int_0^{+\infty} F_c(\omega) \cos \omega x d\omega, \qquad (13)$$

:де

$$F_c(\omega) = \sqrt{\frac{2}{\pi}} \int_0^{+\infty} f(t) \cos \omega t dt \qquad (14)$$

$$f(x) = \sqrt{\frac{2}{\pi}} \int_0^{+\infty} F_s(\omega) \sin \omega x d\omega, \qquad (15)$$

де

$$F_s(\omega) = \sqrt{\frac{2}{\pi}} \int_0^{+\infty} f(t) \sin \omega t dt. \qquad (16)$$

Функции $F_c(\omega)$ и $F_s(\omega)$ называются соответственно косинус-преобразованием и синус-преобразованием для данной функции $f(x)$.

Если равенство (13) рассматривать как интегральное уравнение для функции $F_c(\omega)$, то формула (14) дает его решение.

Аналогично решение интегрального уравнения (15) для функции $F_s(\omega)$ дается формулой (16).

Установим связь между преобразованием Фурье, косинус-преобразованием и синус-преобразованием. Имеем:

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t) \cdot e^{i\omega t} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t) (\cos \omega t + i \sin \omega t) dt =$$

$$= \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{0} f(t) \cos \omega t dt + \int_0^{+\infty} f(t) \cos \omega t dt \right) +$$

$$+ \frac{i}{\sqrt{2\pi}} \left(\int_{-\infty}^{0} f(t) \sin \omega t dt + \int_0^{+\infty} f(t) \sin \omega t dt \right).$$

Если в интегралах $\int_{-\infty}^{0} f(t) \cos \omega t dt$ и $\int_{-\infty}^{0} f(t) \sin \omega t dt$ вместо переменной интегрирования t взять $-t$, то получим

$$F(\omega) = \sqrt{\frac{2}{\pi}} \int_0^{+\infty} \frac{f(t) + f(-t)}{2} \cos \omega t dt +$$

$$+ i \sqrt{\frac{2}{\pi}} \int_0^{+\infty} \frac{f(t) - f(-t)}{2} \sin \omega t dt.$$

Введем обозначение

$$g(x) = \frac{f(x) + f(-x)}{2} \tag{17}$$

и

$$h(x) = \frac{f(x) - f(-x)}{2}, \tag{18}$$

тогда

$$F(\omega) = G_c(\omega) + iH_s(\omega), \tag{19}$$

где $G_c(\omega)$ есть косинус-преобразование для функции $g(x)$, а $H_s(\omega)$ есть синус-преобразование для функции $h(x)$.

Формула (19) устанавливает связь между преобразованием Фурье, косинус-преобразованием и синус-преобразованием.

§ 11. ПРИМЕРЫ РАЗЛОЖЕНИЯ ФУНКЦИЙ В ИНТЕГРАЛ ФУРЬЕ

1. Разложить в интеграл Фурье функцию

$$f(x) = \begin{cases} h \text{ при } \alpha \leqslant x \leqslant \alpha + \Delta, \\ 0 \text{ при } x < \alpha \text{ и } x > \alpha + \Delta, \end{cases}$$

график которой изображен на рис. 21 (прямоугольный импульс).

Р е ш е н и е. Очевидно, эта функция абсолютно интегрируема и кусочно-гладкая в $]-\infty; +\infty[$, поэтому интеграл Фурье будет сходиться при всех x. Найдем коэффициенты $A(\omega)$ и $B(\omega)$.

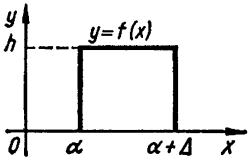

Рис. 21

$$A(\omega) = \frac{1}{\pi}\int_{-\infty}^{+\infty} f(t)\cos\omega t dt = \frac{h}{\pi}\int_{\alpha}^{\alpha+\Delta}\cos\omega t dt =$$

$$= \frac{2h}{\pi\omega}\cos\omega\left(\alpha + \frac{\Delta}{2}\right)\sin\frac{\omega\Delta}{2},$$

$$B(\omega) = \frac{1}{\pi}\int_{-\infty}^{+\infty} f(t)\sin\omega t dt = \frac{h}{\pi}\int_{\alpha}^{\alpha+\Delta}\sin\omega t dt =$$

$$= \frac{2h}{\pi\omega}\sin\omega\left(\alpha + \frac{\Delta}{2}\right)\sin\frac{\omega\Delta}{2}.$$

Интеграл Фурье для заданной функции $f(x)$ примет вид

$$\int_0^{+\infty}(A(\omega)\cos\omega x + B(\omega)\sin\omega x)\,d\omega = \frac{2h}{\pi}\int_0^{+\infty}\left(\cos\omega x\cos\omega\left(\alpha + \frac{\Delta}{2}\right) + \right.$$

$$\left. + \sin\omega x\sin\omega\left(\alpha + \frac{\Delta}{2}\right)\right)\frac{\sin\frac{\omega\Delta}{2}}{\omega}d\omega = \frac{2h}{\pi}\int_0^{+\infty}\frac{\sin\frac{\omega\Delta}{2}}{\omega}\cos\left(x - \alpha - \right.$$

$$\left. - \frac{\Delta}{2}\right)d\omega. \tag{20}$$

Несобственный интеграл (20) сходится при всех x и совпадает с функцией $f(x)$ во всех точках ее непрерывности. В точках $x_1 = \alpha$ и $x_2 = \alpha + \Delta$ значение интеграла равно $\frac{h}{2}$.

2. Разложить в интеграл Фурье функцию

$$f(x)=\begin{cases} h\left(1-\dfrac{2}{\Delta}|x|\right) & \text{при } |x|\leqslant\dfrac{\Delta}{2}, \\ 0 \text{ при } |x|>\dfrac{\Delta}{2}, \end{cases}$$

график которой изображен на рис. 22 (треугольный импульс).

Р е ш е н и е. Функция $f(x)$ четная, поэтому можно применить формулу (13). Найдем косинус-преобразование $f(x)$:

$$F_c(\omega)=\sqrt{\frac{2}{\pi}}\int_0^{+\infty} f(t)\cos\omega t dt = h\sqrt{\frac{2}{\pi}}\times$$

$$\times\int_0^{\frac{\Delta}{2}}\left(1-\frac{2t}{\Delta}\right)\cos\omega t dt.$$

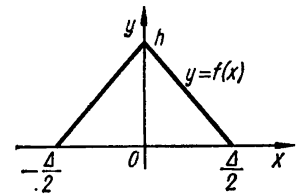

Рис. 22

Интегрируя по частям, получим

$$F_c(\omega)=h\sqrt{\frac{2}{\pi}}\left(\frac{\sin\frac{\omega\Delta}{4}}{\frac{\omega\Delta}{4}}\right)^2\frac{\Delta}{4},$$

тогда

$$f(x)=\sqrt{\frac{2}{\pi}}\int_0^{+\infty}F_c(\omega)\cos\omega x d\omega=\frac{h\Delta}{2\pi}\int_0^{+\infty}\left(\frac{\sin\frac{\omega\Delta}{4}}{\frac{\omega\Delta}{4}}\right)^2\cos\omega x d\omega.$$

Так как функция $f(x)$ непрерывна, то интеграл Фурье при всех x совпадает с $f(x)$.

3. Разложить в интеграл Фурье функцию

$$f(x)=e^{-\alpha x}\quad(\alpha>0,\ x\geqslant 0).$$

Р е ш е н и е. Здесь можно воспользоваться либо формулой (6), либо формулой (7), предварительно доопределив $f(x)$ в $]-\infty;\ 0[$ соответственно четным или нечетным образом. Найдем

$$F_c(\omega)=\sqrt{\frac{2}{\pi}}\int_0^{+\infty}e^{-\alpha t}\cos\omega t dt.$$

Применяя известную формулу для $\int e^{ax}\cos bx dx$, получим

$$F_c(\omega)=\sqrt{\frac{2}{\pi}}\cdot\frac{e^{-\alpha t}}{\alpha^2+\omega^2}(\omega\sin\omega t-\alpha\cos\omega t)\Big|_0^{+\infty}=\sqrt{\frac{2}{\pi}}\cdot\frac{\alpha}{\alpha^2+\omega^2}. \tag{21}$$

Тогда по формуле (13) найдем

$$e^{-\alpha x}=\sqrt{\frac{2}{\pi}}\int_0^{+\infty}F_c(\omega)\cos\omega x d\omega=\frac{2\alpha}{\pi}\int_0^{+\infty}\frac{\cos\omega x}{\alpha^2+\omega^2}d\omega\quad(0<x<+\infty). \tag{22}$$

Вычислим $F_s(\omega) = \sqrt{\frac{2}{\pi}} \int_0^{+\infty} e^{-\alpha t} \sin \omega t dt$. Применяя формулу для $\int e^{\alpha x} \sin bx dx$, найдем

$$F_s(\omega) = \sqrt{\frac{2}{\pi}} \cdot \frac{\omega}{\alpha^2 + \omega^2}. \tag{23}$$

Тогда по формуле (15) имеем

$$e^{-\alpha x} = \frac{2}{\pi} \int_0^{+\infty} \frac{\omega \sin \omega x}{\alpha^2 + \omega^2} d\omega \quad (0 < x < +\infty). \tag{24}$$

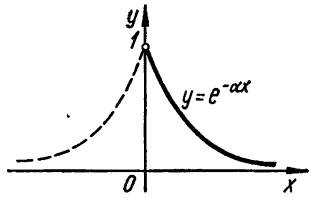

Рис. 23

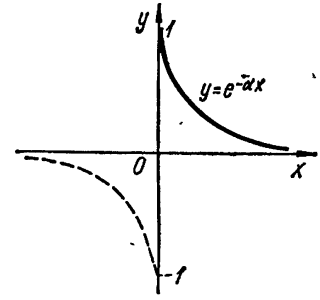

Рис. 24

Несобственные интегралы в формулах (22) и (24) сходятся при всех x. Их графики изображены соответственно на рис. 23 и 24.

Как видим, интеграл (22) при $x = 0$ равен 1, а интеграл (24) при $x = 0$ равен 0.

4. Написать интеграл Фурье в комплексной форме для функции

$$f(x) = \begin{cases} e^{-\alpha x} \ (\alpha > 0) & \text{при } x > 0, \\ 0 & \text{при } x \leqslant 0. \end{cases}$$

Р е ш е н и е. Эту задачу можно решить путем применения формулы (9) или формул (11) и (12). Можно поступить и следующим образом. По формулам (17) и (18) найдем функции:

$$g(x) = \frac{f(x) + f(-x)}{2} = \frac{e^{-\alpha x} + 0}{2} = \frac{1}{2} e^{-\alpha x} \qquad (x > 0),$$

$$h(x) = \frac{f(x) - f(-x)}{2} = \frac{e^{-\alpha x} - 0}{2} = \frac{1}{2} e^{-\alpha x} \qquad (x > 0).$$

Используя формулы (21) и (23), полученные при решении предыдущего примера, найдем $G_c(\omega)$ (косинус-преобразование функции $g(x)$) и $H_s(\omega)$ (синус-преобразование функции $h(x)$):

$$G_c(\omega) = \frac{1}{2} \sqrt{\frac{2}{\pi}} \cdot \frac{\alpha}{\alpha^2 + \omega^2},$$

$$H_s(\omega) = \frac{1}{2} \sqrt{\frac{2}{\pi}} \cdot \frac{\omega}{\alpha^2 + \omega^2}.$$

По формуле (19) получим преобразование Фурье для заданной функции $f(x)$:

$$F(\omega) = G_c(\omega) + iH_s(\omega) = \frac{\alpha + i\omega}{\sqrt{2\pi}\,(\alpha^2 + \omega^2)}.$$

По формуле (12) интеграл Фурье в комплексной форме для функции $f(x)$ будет иметь вид

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\omega x}\,\frac{\alpha+i\omega}{\alpha^2+\omega^2}\,d\omega.$$

5. Разложить в интеграл Фурье на положительной полуоси функцию

$$f(x)=\begin{cases}\sin x & \text{при } 0<x<\pi,\\ 0 & \text{при } x\geqslant\pi,\end{cases}$$

график которой изображен на рис. 25 (полуволна синусоиды).

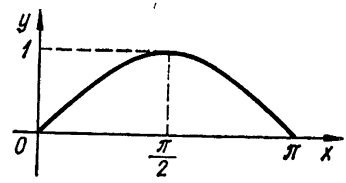

Рис. 25

Р е ш е н и е. Как и в примере 3, определим значения $F_c(\omega)$ и $F_s(\omega)$:

$$F_c(\omega)=\sqrt{\frac{2}{\pi}}\int_0^{\pi}\sin t\cos\omega t\,dt=\frac{1}{\sqrt{2\pi}}\int_0^{\pi}(\sin(1+\omega)t+\sin(1-\omega)t)\,dt=$$

$$=\frac{1}{\sqrt{2\pi}}\left(-\frac{\cos(1+\omega)t}{1+\omega}-\frac{\cos(1-\omega)t}{1-\omega}\right)\Bigg|_0^{\pi}=$$

$$=\frac{1}{\sqrt{2\pi}}\left(\frac{2}{1-\omega^2}-\frac{\cos(1+\omega)\pi}{1+\omega}-\frac{\cos(1-\omega)\pi}{1-\omega}\right)=$$

$$=\frac{2}{\sqrt{2\pi}}\cdot\frac{1+\cos\omega\pi}{1-\omega^2}=\sqrt{\frac{2}{\pi}}\cdot\frac{1+\cos\omega\pi}{1-\omega^2}.$$

$$F_s(\omega)=\sqrt{\frac{2}{\pi}}\int_0^{\pi}\sin t\sin\omega t\,dt=\sqrt{\frac{2}{\pi}}\cdot\frac{\sin\omega\pi}{1-\omega^2}.$$

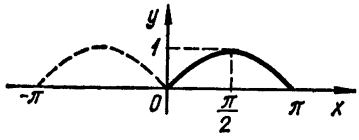

Рис. 26

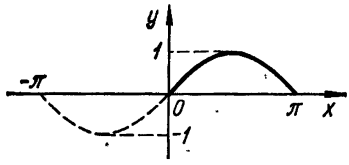

Рис. 27

Тогда

$$f(x)=\sqrt{\frac{2}{\pi}}\int_0^{+\infty}F_c(\omega)\cos\omega x\,d\omega=\frac{2}{\pi}\int_0^{+\infty}\frac{1+\cos\omega\pi}{1-\omega^2}\cos\omega x\,d\omega\quad(x\geqslant 0)\quad(25)$$

или

$$f(x)=\sqrt{\frac{2}{\pi}}\int_0^{+\infty}F_s(\omega)\sin\omega x\,d\omega=\frac{2}{\pi}\int_0^{+\infty}\frac{\sin\omega\pi}{1-\omega^2}\sin\omega x\,d\omega\quad(x\geqslant 0).\quad(26)$$

Интегралы (25) и (26) сходятся и при отрицательных x. Их графики изображены соответственно на рис. 26 и 27.

6. Разложить в интеграл Фурье функцию

$$f(x)=e^{-\frac{x^2}{2\sigma^2}},$$

график которой изображен на рис. 28. (Эта функция часто встречается в радиотехнике. Сигнал такой формы называется колокольным импульсом).

Решение. Найдем

$$F_c(\omega) = \sqrt{\frac{2}{\pi}} \int_0^{+\infty} e^{-\frac{t^2}{2\sigma^2}} \cos \omega t dt = \sigma \sqrt{\frac{2}{\pi}} \int_0^{+\infty} e^{-\frac{\tau^2}{2}} \cos \omega\sigma\tau d\tau$$

(здесь выполнена замена переменной $\frac{t}{\sigma} = \tau$). Запишем

$$F_c(\omega) = \sigma \sqrt{\frac{2}{\pi}} \int_0^{+\infty} e^{-\frac{t^2}{2}} \cos \omega\sigma t dt.$$

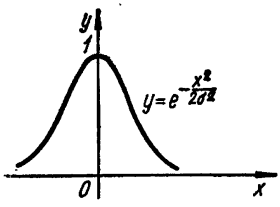

Рис. 28

Рассмотрим несобственный интеграл, зависящий от параметра λ:

$$S(\lambda) = \int_0^{+\infty} e^{-\frac{t^2}{2}} \cos \lambda t dt. \qquad (27)$$

Можно проверить, что интеграл (27) имеет производную по λ, причем,

$$S'(\lambda) = \int_0^{+\infty} \frac{d}{d\lambda} \left(e^{-\frac{t^2}{2}} \cos \lambda t\right) dt = -\int_0^{+\infty} e^{-\frac{t^2}{2}} t \sin \lambda t dt.$$

Выполнив интегрирование по частям, получим

$$S'(\lambda) = -\left(-\sin \lambda t \cdot e^{-\frac{t^2}{2}}\Big|_0^{+\infty} + \lambda \int_0^{+\infty} e^{-\frac{t^2}{2}} \cos \lambda t dt\right),$$

откуда

$$S'(\lambda) = -\lambda S(\lambda).$$

Это дифференциальное уравнение первого порядка с разделяющимися переменными. Решая его, получим

$$S(\lambda) = Ce^{-\frac{\lambda^2}{2}}. \qquad (28)$$

Положив $\lambda = \omega\sigma$, имеем

$$F_c(\omega) = \sigma \sqrt{\frac{2}{\pi}} Ce^{-\frac{\omega^2\sigma^2}{2}},$$

тогда

$$e^{-\frac{x^2}{2\sigma^2}} = \sqrt{\frac{2}{\pi}} \int_0^{+\infty} F_c(\omega) \cos \omega x d\omega = C \frac{2\sigma}{\pi} \int_0^{+\infty} e^{-\frac{\omega^2\sigma^2}{2}} \cos \omega x d\omega. \qquad (29)$$

Остается определить лишь неизвестную постоянную C. Положим в интеграле (29) $\omega = \frac{\omega_1}{\sigma}$. Получим

$$e^{-\frac{x^2}{2\sigma^2}} = \frac{2}{\pi} C \int_0^{+\infty} e^{-\frac{\omega_1^2}{2}} \cos \frac{\omega_1 x}{\sigma} d\omega_1.$$

Принимая $\omega_1 = t$, найдем

$$e^{-\frac{x^2}{2\sigma^2}} = \frac{2}{\pi} C \int_0^{+\infty} e^{-\frac{t^2}{2}} \cos \frac{x}{\sigma} t dt.$$

Принимая во внимание равенство (27), запишем последнюю формулу так:

$$e^{-\frac{x^2}{2\sigma^2}} = \frac{2}{\pi} C \cdot S\left(\frac{x}{\sigma}\right),$$

но из равенства (28)

$$S\left(\frac{x}{\sigma}\right) = Ce^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} = Ce^{-\frac{x^2}{2\sigma^2}},$$

тогда

$$e^{-\frac{x^2}{2\sigma^2}} = \frac{2}{\pi} C^2 \cdot e^{-\frac{x^2}{2\sigma^2}},$$

откуда

$$C = \sqrt{\frac{\pi}{2}}.$$

Следовательно,

$$e^{-\frac{x^2}{2\sigma^2}} = \sqrt{\frac{2}{\pi}}\, \sigma \int_0^{+\infty} e^{-\frac{\omega^2\sigma^2}{2}} \cos \omega x d\omega. \qquad (30)$$

Это интеграл Фурье для колокольного импульса.

Отметим, что косинус-преобразование Фурье функции $e^{-\frac{x^2}{2\sigma^2}}$ является функцией того же вида $\sigma e^{-\frac{\omega^2\sigma^2}{2}}$.

В электро- и радиотехнической литературе интеграл Фурье в комплексной форме часто записывают так:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S(j\omega) e^{j\omega t} d\omega,$$

где комплексная величина

$$S(j\omega) = \int_{-\infty}^{+\infty} f(t) e^{-j\omega t} dt \quad (j^2 = -1)$$

называется спектральной плотностью или спектральной характеристикой, а ее модуль $S(\omega)$ — спектром функции (сигнала, временной функции) $f(t)$. При этом

$$S(j\omega) = S(\omega) \cdot e^{-j\alpha(\omega)} = a(\omega) - jb(\omega),$$

$$a(\omega) = \int_{-\infty}^{+\infty} f(t) \cos \omega t dt,$$

$$b(\omega) = \int_{-\infty}^{+\infty} f(t) \sin \omega t dt,$$

$$S(\omega) = \sqrt{(a(\omega))^2 + (b(\omega))^2},$$

$$\alpha(\omega) = \operatorname{arctg} \frac{b(\omega)}{a(\omega)}.$$

Модуль спектральной плотности $S(\omega)$ — функция четная, фаза спектральной плотности $\alpha(\omega)$ — нечетная функция частоты ω.

7. Найти спектральную плотность, модуль и фазу спектральной плотности экспоненциального импульса

$$f(t) = \begin{cases} Be^{-\beta t} & \text{при } t \geqslant 0, \\ 0 & \text{при } t < 0. \end{cases}$$

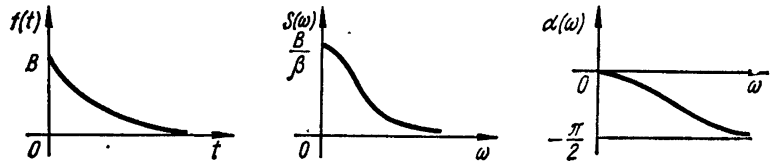

Рис. 29

Решение.

$$S(j\omega) = \int_{-\infty}^{+\infty} f(t)\, e^{-j\omega t} dt = \int_{0}^{+\infty} Be^{-(\beta + j\omega)t} dt = \frac{B}{\beta + j\omega} = \frac{B(\beta - j\omega)}{\beta^2 + \omega^2};$$

$$a(\omega) = \frac{B\beta}{\beta^2 + \omega^2}; \quad b(\omega) = -\frac{B\omega}{\beta^2 + \omega^2};$$

$$S(\omega) = \sqrt{(a(\omega))^2 + (b(\omega))^2} = \frac{B}{\sqrt{\beta^2 + \omega^2}};$$

$$\alpha(\omega) = \operatorname{arctg} \frac{b(\omega)}{a(\omega)} = -\operatorname{arctg} \frac{\omega}{\beta}.$$

Графики $f(t)$, $S(\omega)$ и $\alpha(\omega)$ изображены на рис. 29.

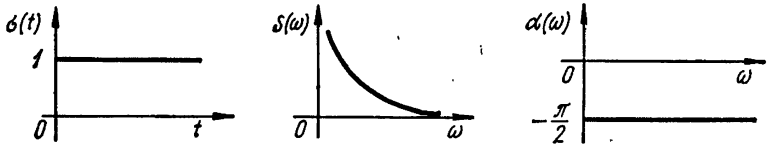

Рис. 30

8. Найти спектральную плотность, модуль и фазу спектральной плотности единичной функции

$$\sigma(t) = \begin{cases} 1 & \text{при } t \geqslant 0, \\ 0 & \text{при } t < 0. \end{cases}$$

Решение. Единичная функция не будет абсолютно интегрируемой, так как

$$\int_{-\infty}^{+\infty} |\sigma(t)|\, dt = \int_{0}^{+\infty} dt = +\infty.$$

Поэтому представить ее интегралом Фурье нельзя. Однако ее спектральную плотность можно рассматривать как предел спектральной плотности экспоненци-

ального импульса, рассмотренного в примере 7, когда $\beta \to 0$, $B = 1$. Получим

$$S(j\omega) = \lim_{\beta \to 0} \frac{1}{\beta + j\omega} = \frac{1}{j\omega} = -\frac{j}{\omega} = \frac{1}{\omega} e^{-j\frac{\pi}{2}},$$

$$S(\omega) = \frac{1}{\omega}, \quad \alpha(\omega) = -\frac{\pi}{2}.$$

Графики функций $\sigma(t)$, $S(\omega)$, $\alpha(\omega)$ изображены на рис. 30.

9. Найти спектральную плотность и модуль спектральной плотности прямоугольного импульса

$$f(t) = \begin{cases} B & \text{при } 0 \leqslant t \leqslant \tau, \\ 0 & \text{при } t < 0 \text{ и при } t > \tau. \end{cases}$$

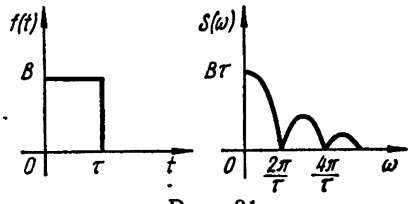

Рис. 31

Решение.

$$S(j\omega) = \int_0^\tau Be^{-j\omega t}dt = -\frac{B}{j\omega} e^{-j\omega t}\Big|_0^\tau = \frac{B(1 - e^{-j\omega\tau})}{j\omega} = \frac{Bj(e^{-j\omega\tau} - 1)}{\omega} =$$

$$= \frac{Bj}{\omega}(\cos\omega\tau - j\sin\omega\tau - 1) = \frac{B}{\omega}(\sin\omega\tau - j(1 - \cos\omega\tau)),$$

$$S(\omega) = \frac{B}{\omega}\sqrt{\sin^2\omega\tau + (1 - \cos\omega\tau)^2} =$$

$$= \frac{B}{\omega}\sqrt{\sin^2\omega\tau + \cos^2\omega\tau + 1 - 2\cos\omega\tau} = \frac{B}{\omega}\sqrt{2(1 - \cos\omega\tau)} =$$

$$= \frac{2B}{\omega}\left|\sin\frac{\omega\tau}{2}\right|.$$

Графики функций $f(t)$ и $S(\omega)$ изображены на рис. 31.

Преобразование Фурье дает возможность получить удобное и наглядное представление процессов, протекающих в электро- и радиотехнических цепях при наличии в них электрических и магнитных величин, изменяющихся во времени по несинусоидальному непериодическому закону.

ГЛАВА III

УРАВНЕНИЯ МАТЕМАТИЧЕСКОЙ ФИЗИКИ

§ 12. ПОНЯТИЕ ДИФФЕРЕНЦИАЛЬНОГО УРАВНЕНИЯ С ЧАСТНЫМИ ПРОИЗВОДНЫМИ

В курсе высшей математики изучались обыкновенные дифференциальные уравнения, решениями которых являются функции одного аргумента. Однако многие задачи в математике, физике, электро- и радиотехнике и в других науках приводят к дифференциальным уравнениям относительно функций двух, трех и большего числа аргументов.

Если неизвестная функция $z = f(x, y)$ является функцией двух переменных, то в общем случае дифференциальное уравнение с част-

ными производными первого порядка для этой функции имеет вид

$$F\left(x, y, z, \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right) = 0.$$

Уравнение с частными производными второго порядка для той же функции будет

$$F\left(x, y, z, \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}, \frac{\partial^2 z}{\partial x^2}, \frac{\partial^2 z}{\partial x \partial y}, \frac{\partial^2 z}{\partial y^2}\right) = 0.$$

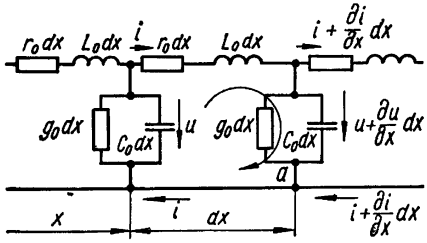

Рис. 32

Аналогично можно написать для этой функции и дифференциальное уравнение порядка n.

В частности, в электротехнике, радиотехнике и особенно в технике электрической связи часто рассматриваются дифференциальные уравнения двухпроводной однородной линии (длинной линии) с частными производными второго порядка, где независимыми переменными являются расстояние x, отсчитываемое от начала или от конца линии, и время t, а неизвестными функциями — мгновенные значения тока $i(x, t)$ и напряжения $u(x, t)$ в любом сечении линии.

В однородной двухпроводной линии сопротивление и индуктивность проводов, а также проводимость и емкость между проводами распределены равномерно вдоль всей ее длины. Поэтому под первичными параметрами однородной линии понимают отнесенные к единице длины: r_0 — активное сопротивление прямого и обратного проводов; L_0 — индуктивность петли, образуемой прямым и обратным проводами; g_0 — проводимость (утечка), обусловленная несовершенством изоляции между проводами; C_0 — емкость между проводами.

Однородные двухпроводные линии в радиотехнике и в технике электропроводной связи выполняются либо в виде параллельных проводов, либо коаксиального кабеля. Уравнения для токов и напряжений таких линий в принципе могут быть применены к трехфазным и другим типам линий электропередачи, применяемым в электроэнергетике. Первичные параметры двухпроводной однородной линии r_0, L_0, g_0 и C_0 зависят от ее конструкции и частоты протекающего в ней переменного тока. Будем считать первичные параметры линии постоянными и с учетом этого получим уравнения длинной линии.

Представим линию в виде множества последовательно соединенных достаточно малых элементов длиной dx, каждый из которых обладает сопротивлением r_0dx, индуктивностью L_0dx, проводимостью g_0dx и емкостью C_0dx (рис. 32). При этом будем считать, что сопротивление r_0dx и индуктивность L_0dx элемента линии включены лишь в один прямой (верхний) провод, а расстояние x отсчи-

тывается от начала до текущего элемента ее длины. Обозначим мгновенные значения тока и напряжения в начале выбранного элемента линии dx через i и u; тогда в конце этого элемента (или в начале следующего) они будут $i + \frac{\partial i}{\partial x}dx$ и $u + \frac{\partial u}{\partial x}dx$. Положительные направления токов и напряжений выберем, как показано на рис. 32.

На основании первого закона Кирхгофа для узла a эквивалентной схемы можно утверждать, что отходящий от узла ток i равен сумме подходящих к этому узлу токов: тока в конце элемента $i + \frac{\partial i}{\partial x}dx$, тока утечки $g_0\left(u + \frac{\partial u}{\partial x}dx\right)dx$ между проводами через проводимость $g_0 dx$ и тока смещения в диэлектрике $C_0 \frac{\partial}{\partial t}\left(u + \frac{\partial u}{\partial x}dx\right)dx$ между проводами через емкость $C_0 dx$, т. е.

$$i = \left(i + \frac{\partial i}{\partial x}dx\right) + g_0\left(u + \frac{\partial u}{\partial x}dx\right)dx + C_0\frac{\partial}{\partial t}\left(u + \frac{\partial u}{\partial x}dx\right)dx. \quad (1)$$

На основании второго закона Кирхгофа для контура, охватывающего выбранный элемент линии dx, получим, что напряжение u в начале этого элемента равно сумме падения напряжения $r_0 i dx$ на активном сопротивлении $r_0 dx$, падения напряжения $L_0\frac{\partial i}{\partial t}dx$ на индуктивности $L_0 dx$ и напряжения $u + \frac{\partial u}{\partial x}dx$ в конце элемента, т. е.

$$u = r_0 i dx + L_0\frac{\partial i}{\partial t}dx + \left(u + \frac{\partial u}{\partial x}dx\right). \quad (2)$$

Приводя подобные члены в уравнениях (1) и (2) и пренебрегая величинами второго порядка малости, после сокращения на dx получаем дифференциальные уравнения однородной двухпроводной линии в частных производных:

$$\begin{cases} -\frac{\partial i}{\partial x} = g_0 u + C_0\frac{\partial u}{\partial t}, & (3) \\ -\frac{\partial u}{\partial x} = r_0 i + L_0\frac{\partial i}{\partial t}. & (4) \end{cases}$$

Эти уравнения иногда называют телеграфными уравнениями. Они остаются справедливыми при любых изменениях тока и напряжения во времени.

Полученную систему уравнений (3) и (4) можно привести к уравнению, содержащему только искомый ток $i(x, t)$ или только искомое напряжение $u(x, t)$.

Действительно, если продифференцировать все члены уравнения (3) по переменной x, а все члены уравнения (4) по переменной t и умножить их на C_0, то после вычитания полученных выражений будем иметь

$$-\frac{\partial^2 i}{\partial x^2} = g_0\frac{\partial u}{\partial x} - r_0 C_0\frac{\partial i}{\partial t} - L_0 C_0\frac{\partial^2 i}{\partial t^2}.$$

Подставив в это уравнение выражение $\frac{\partial u}{\partial x}$ из уравнения (4), найдем

$$-\frac{\partial^2 i}{\partial x^2} = -g_0\left(r_0 i + L_0\frac{\partial i}{\partial t}\right) - r_0 C_0\frac{\partial i}{\partial t} - L_0 C_0\frac{\partial^2 i}{\partial t^2},$$

или

$$\frac{\partial^2 i}{\partial x^2} - L_0 C_0\frac{\partial^2 i}{\partial t^2} - (r_0 C_0 + g_0 L_0)\frac{\partial i}{\partial t} - r_0 g_0 i = 0. \qquad (5)$$

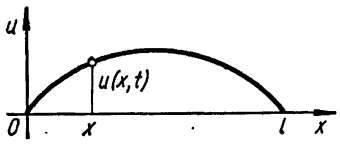

Рис. 33

Аналогично получается и уравнение для $u(x, t)$:

$$\frac{\partial^2 u}{\partial x^2} - L_0 C_0\frac{\partial^2 u}{\partial t^2} - (r_0 C_0 + g_0 L_0)\frac{\partial u}{\partial t} - r_0 g_0 u = 0. \qquad (6)$$

При расчете электростатических полей, магнитных полей постоянных токов, а также при рассмотрении задач о стационарном тепловом режиме, задач диффузии, гидро- и аэродинамики встречается уравнение Лапласа

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \qquad (7)$$

или уравнение Пуассона

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + a = 0. \qquad (8)$$

В механике исследуется волновое уравнение — уравнение колебаний струны, под которой понимают гибкую упругую нить, закрепленную в точках $x = 0$ и $x = l$ (рис. 33). Уравнение движения струны имеет вид

$$\frac{\partial^2 u}{\partial t^2} = a^2\frac{\partial^2 u}{\partial x^2}. \qquad (9)$$

Задача интегрирования уравнения с частными производными состоит в том, чтобы найти все решения данного уравнения.

Ясно, что изученные ранее обыкновенные дифференциальные уравнения — это простейший случай дифференциальных уравнений с частными производными, когда число аргументов неизвестной функции равно единице. Вместе с тем, известно, что обыкновенное дифференциальное уравнение имеет бесчисленное множество решений, так как его общее решение содержит произвольные постоянные. Естественно предположить, что и уравнение с частными производными имеет бесконечное множество решений. Оказывается, что в решение уравнения с частными производными входят даже произвольные функции тех же аргументов, от которых зависит неизвестная функция. Покажем это на примерах.

Примеры

1. Решить уравнение с частными производными первого порядка

$$\frac{\partial z}{\partial x} - 2xy = 0 \quad (z = z(x, y)).$$

Р е ш е н и е. Запишем данное уравнение в виде $\frac{\partial z}{\partial x} = 2xy$.

Будем рассматривать y как параметр, тогда данное уравнение становится обыкновенным дифференциальным уравнением для функции z аргумента x. Его общее решение имеет вид $z = x^2y + C$.

Если теперь вместо произвольной постоянной C взять произвольную функцию $\varphi(y)$, то получим общее решение данного дифференциального уравнения с частными производными $z = x^2y + \varphi(y)$.

При этом, какой бы ни была произвольная функция $\varphi(y)$, функция $z = x^2y + \varphi(y)$ удовлетворяет данному дифференциальному уравнению. Действительно,

$$\frac{\partial z}{\partial x} = \frac{\partial}{\partial x}(x^2y + \varphi(y)) = 2xy + \frac{\partial}{\partial x}\varphi(y) = 2xy,$$

так как $\frac{\partial}{\partial x}\varphi(y) = 0$.

Из рассмотренного примера видно, что общее решение уравнения с частными производными первого порядка всегда содержит одну произвольную функцию.

2. Решить уравнение $\frac{\partial^2 z}{\partial x \partial y} = 0$.

Р е ш е н и е. Запишем это уравнение с частными производными второго порядка так:

$$\frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right) = 0.$$

Отсюда следует, что $\frac{\partial z}{\partial y}$ не зависит от x, т. е. $\frac{\partial z}{\partial y} = \theta(y)$, где $\theta(y)$ — любая функция y.

Интегрируя последнее равенство по y, найдем

$$z = \int \theta(y)\,dy + \varphi(x),$$

где $\varphi(x)$ — произвольная функция, зависящая только от x. (Эта функция играет роль произвольной постоянной при выполнении интегрирования по y.)

Итак,

$$z = \varphi(x) + \psi(y),$$

где $\psi(y) = \int \theta(y)\,dy$ — произвольная функция y, а $\varphi(x)$ — указанная выше произвольная функция x.

В рассмотренном примере оказалось, что решение уравнения с частными производными второго порядка зависит от двух произвольных функций (конечно, дифференцируемых).

Следует заметить, что полученные в этих примерах результаты нельзя распространить на общий случай дифференциального уравнения с частными производными порядка n.

Напомним, что с помощью произвольных постоянных, входящих в общее решение обыкновенного дифференциального уравнения, можно найти его частное решение, удовлетворяющее начальным условиям (решить задачу Коши для этого уравнения). Наличие

произвольных функций в решении уравнения с частными производными также используется для того, чтобы найти частное решение этого уравнения, удовлетворяющее некоторым условиям, называемым краевыми условиями.

К краевым условиям относятся: начальные условия, описывающие состояние системы в начальный момент времени (при $t = 0$), и граничные условия, которым должна удовлетворять искомая функция $u(x, t)$, например, при $x = 0$ и $x = l$. Краевые условия формулируются всегда из физических условий задачи. Так, например, для однородной двухпроводной линии начальными условиями являются мгновенные значения тока и напряжения в начале (или в конце) линии при $t = 0$, а граничные условия будут определяться связями между токами и напряжениями в начале (или в конце) линии, которые в свою очередь зависят от заданного режима линии. Для струны, концы которой закреплены неподвижно (при $x = 0$ и $x = l$) и форма которой в начальный момент времени ($t = 0$) задается функцией $\varphi(x)$, начальными условиями будут равенства:

$$u(x, 0) = u|_{t=0} = \varphi(x) \tag{10}$$

и

$$u_t(x, 0) = \left.\frac{\partial u}{\partial t}\right|_{t=0} = \psi(x), \tag{11}$$

где $u_t(x, 0)$ — скорость в каждой точке струны при $t = 0$, заданная функцией $\psi(x)$, а граничными условиями являются равенства:

$$u(0, t) = 0 \tag{12}$$

и

$$u(l, t) = 0. \tag{13}$$

Обычно в кратком курсе не изучается теория интегрирования дифференциальных уравнений с частными производными первого порядка. Поэтому ниже рассматриваются лишь некоторые основные вопросы, относящиеся к решению отдельных видов дифференциальных уравнений с частными производными второго порядка, наиболее часто встречающихся в практике.

§ 13. КЛАССИФИКАЦИЯ ЛИНЕЙНЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ С ЧАСТНЫМИ ПРОИЗВОДНЫМИ ВТОРОГО ПОРЯДКА

Условимся неизвестную функцию обозначать буквой u и предположим, что u зависит от двух переменных x, y (или x, t). Введем обозначения:

$$\frac{\partial u}{\partial x} = u_x, \quad \frac{\partial u}{\partial y} = u_y, \quad \frac{\partial^2 u}{\partial x^2} = u_{xx}, \quad \frac{\partial^2 u}{\partial x \partial y} = u_{xy}, \quad \frac{\partial^2 u}{\partial y^2} = u_{yy}.$$

Рассмотрим дифференциальное уравнение с частными производными второго порядка

$$a_{11}u_{xx} + 2a_{12}u_{xy} + a_{22}u_{yy} + b_1u_x + b_2u_y + cu = f(x, y),$$

где a_{11}, a_{12}, a_{22}, b_1, b_2, c, $f(x, y)$ — заданные функции x, y.

Это дифференциальное уравнение называется линейным, так как оно линейно относительно функции u, ее первых и вторых частных производных. Если $f(x, y) \equiv 0$, то уравнение называется линейным однородным, в противном случае — линейным неоднородным (с правой частью или неоднородностью $f(x, y)$).

Если все коэффициенты a_{11}, a_{12}, a_{22}, b_1, b_2 и c — постоянные, то уравнение называется линейным уравнением с постоянными коэффициентами, множитель 2 у коэффициента a_{12} при смешанной производной u_{xy} записывается лишь для удобства.

Можно убедиться в том, что с помощью преобразования переменных $\xi = \varphi(x, y)$, $\eta = \psi(x, y)$ данное дифференциальное уравнение остается линейным и будет иметь вид

$$\bar{a}_{11}u_{\xi\xi} + 2\bar{a}_{12}u_{\xi\eta} + \bar{a}_{22}u_{\eta\eta} + \bar{b}_1u_\xi + \bar{b}_2u_\eta + \bar{c}u = \bar{f}(\xi, \eta),$$

где коэффициенты при производных находятся по формулам:

$$\bar{a}_{11} = a_{11}\xi_x^2 + 2a_{12}\xi_x\xi_y + a_{22}\xi_y^2,$$

$$\bar{a}_{12} = a_{11}\xi_x\eta_x + a_{12}(\xi_x\eta_y + \eta_x\xi_y) + a_{22}\xi_y\eta_y,$$

$$\bar{a}_{22} = a_{11}\eta_x^2 + 2a_{12}\eta_x\eta_y + a_{22}\eta_y^2,$$

$$\bar{b}_1 = a_{11}\xi_{xx} + 2a_{12}\xi_{xy} + a_{22}\xi_{yy} + b_1\xi_x + b_2\xi_y,$$

$$\bar{b}_2 = a_{11}\eta_{xx} + 2a_{12}\eta_{xy} + a_{22}\eta_{yy} + b_1\eta_x + b_2\eta_y.$$

Возникает вопрос: нельзя ли выбрать новые переменные ξ и η так, чтобы в преобразованном уравнении некоторые коэффициенты при вторых производных были равны нулю? Оказывается, что эта задача связана с решением такого обыкновенного дифференциального уравнения:

$$a_{11}dy^2 - 2a_{12}dxdy + a_{22}dx^2 = 0.$$

Это уравнение называется характеристическим для исходного дифференциального уравнения с частными производными, а его интегралы — характеристиками.

Если $\varphi(x, y) = C_1$ — общий интеграл характеристического уравнения, то положив $\xi = \varphi(x, y)$, мы обратим в нуль коэффициент при $u_{\xi\xi}$. Если $\psi(x, y) = C_2$ является другим общим интегралом характеристического уравнения, линейно-независимым от $\varphi(x, y)$, то полагая $\eta = \psi(x, y)$, мы обратим в нуль также и коэффициент при $u_{\eta\eta}$.

Характеристическое уравнение можно записать и так:

$$\frac{dy}{dx} = \frac{a_{12} \pm \sqrt{a_{12}^2 - a_{11}a_{22}}}{a_{11}}.$$

Если $a_{12}^2 - a_{11}a_{22} > 0$, то интегралы характеристического уравнения $\varphi(x, y) = C_1$ и $\psi(x, y) = C_2$ действительны и различны.

Полагая $\xi = \varphi(x, y)$, $\eta = \psi(x, y)$, приведем уравнение с частными производными к виду

$$u_{\xi\eta} = \Phi(\xi, \eta, u, u_\xi, u_\eta).$$

В этом случае говорят, что уравнение имеет гиперболический тип. Если теперь положить $\alpha = \frac{\xi+\eta}{2}$, $\beta = \frac{\xi-\eta}{2}$, то уравнение примет вид

$$u_{\alpha\alpha} - u_{\beta\beta} = \Phi_1(\alpha, \beta, u, u_\alpha, u_\beta).$$

Если $a_{12}^2 - a_{11}a_{22} = 0$, то характеристическое уравнение имеет лишь один общий интеграл $\varphi(x, y) = C$.

Пусть $\eta(x, y)$ — любая функция, линейно не зависящая от $\varphi(x, y)$.

Тогда полагая $\xi = \varphi(x, y)$, $\eta = \eta(x, y)$, приведем уравнение с частными производными к виду

$$u_{\eta\eta} = \Phi(\xi, \eta, u, u_\xi, u_\eta).$$

В этом случае уравнение имеет параболический тип.

Если $a_{12}^2 - a_{11}a_{22} < 0$, то характеристическое уравнение имеет комплексные (сопряженные) общие интегралы

$$\varphi(x, y) = C_1 \quad \text{и} \quad \varphi^*(x, y) = C_2.$$

Полагая $\alpha = \frac{\varphi+\varphi^*}{2}$, $\beta = \frac{\varphi-\varphi^*}{2i}$, уравнение с частными производными приводим к виду

$$u_{\alpha\alpha} + u_{\beta\beta} = \Phi(\alpha, \beta, u, u_\alpha, u_\beta).$$

В этом случае уравнение имеет эллиптический тип.

Если коэффициенты линейного уравнения постоянны, то характеристическое уравнение имеет решения:

$$y = \frac{a_{12} + \sqrt{a_{12}^2 - a_{11}a_{22}}}{a_{11}} x + C_3,$$

$$y = \frac{a_{12} - \sqrt{a_{12}^2 - a_{11}a_{22}}}{a_{12}} x + C_4.$$

При $a_{12}^2 - a_{11}a_{22} > 0$ уравнение приводится к виду

$$\left.\begin{array}{l} u_{\xi\eta} + b_1u_\xi + b_2u_\eta + cu = f \\ \text{или} \\ u_{\xi\xi} - u_{\eta\eta} + b_1u_\xi + b_2u_\eta + cu = f \end{array}\right\} \text{(гиперболический тип).}$$

При $a_{12}^2 - a_{11}a_{22} = 0$ уравнение приводится к виду

$$u_{\xi\xi} + b_1u_\xi + b_2u_\eta = f \text{ (параболический тип).}$$

При $a_{12}^2 - a_{11}a_{22} < 0$ уравнение приводится к виду

$$u_{\xi\xi} + u_{\eta\eta} + b_1u_\xi + b_2u_\eta + cu = f \text{ (эллиптический тип).}$$

Если вместо функции u ввести новую функцию v, положив

$$u = ve^{\lambda\xi+\mu\eta},$$

где λ и μ — постоянные, то при соответствующем выборе значений λ и μ получим следующие канонические формы для уравнений с постоянными коэффициентами:

$$\left.\begin{aligned} v_{\xi\eta} + \gamma v &= f_1 \\ v_{\xi\xi} - v_{\eta\eta} &= f_1 \end{aligned}\right\} \text{ (гиперболический тип),}$$

или

$$v_{\xi\xi} + bv_\eta = f_1 \text{ (параболический тип),}$$

$$v_{\xi\xi} + v_{\eta\eta} + \gamma v = f_1 \text{ (эллиптический тип).}$$

Примеры

1. Найти области гиперболичности, эллиптичности и параболичности уравнения $u_{xx} + yu_{yy} = 0$.

Р е ш е н и е. В данном случае

$$a_{11} = 1,\ a_{12} = 0,\ a_{22} = y \text{ и } a_{12}^2 - a_{11}a_{22} = -y.$$

При $y < 0$ получим $a_{12}^2 - a_{11}a_{22} > 0$. Следовательно, уравнение имеет гиперболический тип.

При $y = 0$ получаем $a_{12}^2 - a_{11}a_{22} = 0$; уравнение имеет параболический тип.

При $y > 0$ получаем $a_{12}^2 - a_{11}a_{22} < 0$; уравнение имеет эллиптический тип.

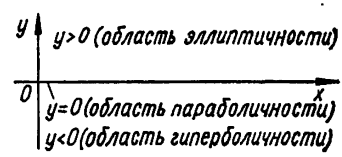

Рис. 34

Эти области показаны на рис. 34.

2. Привести к каноническому виду уравнение

$$u_{xx} + 2u_{xy} + u_{yy} + 4u = 0.$$

Р е ш е н и е. Здесь $a_{11} = 1$, $a_{12} = 1$, $a_{22} = 1$, $a_{12}^2 - a_{11}a_{22} = 0$, т. е. задано уравнение параболического типа.

Характеристическое уравнение

$$\frac{dy}{dx} = \frac{a_{12} \pm \sqrt{a_{12}^2 - a_{11}a_{22}}}{a_{11}}$$

имеет вид $\dfrac{dy}{dx} = 1$. Его общий интеграл $x - y = C$.

Положим $\xi = x - y$. В качестве другой новой переменной возьмем, например, $\eta = x + y$ (очевидно, ξ и η независимы).

При этом $\xi_x = 1$, $\xi_y = -1$, $\eta_x = 1$, $\eta_y = 1$. Тогда

$$u_x = u_\xi\xi_x + u_\eta\eta_x = u_\xi + u_\eta,$$

$$u_y = u_\xi\xi_y + u_\eta\eta_y = u_\eta - u_\xi,$$

$$u_{xx} = u_{\xi\xi}\xi_x + u_{\xi\eta}\eta_x + u_{\eta\xi}\xi_x + u_{\eta\eta}\eta_x = u_{\xi\xi} + 2u_{\xi\eta} + u_{\eta\eta},$$

$$u_{xy} = u_{\xi\xi}\xi_y + u_{\xi\eta}\eta_y + u_{\eta\xi}\xi_y + u_{\eta\eta}\eta_y = u_{\eta\eta} - u_{\xi\xi},$$

$$u_{yy} = u_{\eta\xi}\xi_y + u_{\eta\eta}\eta_y - u_{\xi\xi}\xi_y - u_{\xi\eta}\eta_y = u_{\xi\xi} - 2u_{\xi\eta} + u_{\eta\eta}.$$

Подставим u_{xx}, u_{xy}, u_{yy} в заданное уравнение. После простейших преобразований получим

$$u_{\eta\eta} + u = 0.$$

3. Привести к каноническому виду уравнение

$$u_{xx} + 2u_{xy} + 2u_{yy} = 0.$$

Решение. Здесь $a_{11} = 1$, $a_{12} = 1$; $a_{22} = 2$; $a_{12}^2 - a_{11}a_{22} < 0$, т. е. задано уравнение эллиптического типа.

Характеристическое уравнение

$$\frac{dy}{dx} = \frac{a_{12} \pm \sqrt{a_{12}^2 - a_{11}a_{22}}}{a_{11}} = 1 \pm i,$$

$$\frac{dy}{dx} = 1 + i, \quad y = (1 + i)x + C_1, \quad (y - x) - ix = C_1,$$

$$\frac{dy}{dx} = 1 - i, \quad y = (1 - i)x + C_2, \quad (y - x) + ix = C_2.$$

Введем новые переменные $\xi = (y - x) - ix$ и $\eta = (y - x) + ix$.

Положим $\alpha = \frac{\xi + \eta}{2} = y - x$, $\beta = \frac{\xi - \eta}{-2i} = x$, при этом $\alpha_x = -1$, $\alpha_y = 1$, $\beta_x = 1$, $\beta_y = 0$.

Тогда

$$u_x = u_\alpha \alpha_x + u_\beta \beta_x = -u_\alpha + u_\beta,$$

$$u_y = u_\alpha \alpha_y + u_\beta \beta_y = u_\alpha,$$

$$u_{xx} = -u_{\alpha\alpha}\alpha_x - u_{\alpha\beta}\beta_x + u_{\beta\alpha}\alpha_x + u_{\beta\beta}\beta_x = u_{\alpha\alpha} - 2u_{\alpha\beta} + u_{\beta\beta},$$

$$u_{yy} = u_{\alpha\alpha}\alpha_y + u_{\alpha\beta}\beta_y = u_{\alpha\alpha},$$

$$u_{xy} = -u_{\alpha\alpha}\alpha_y - u_{\alpha\beta}\beta_y + u_{\beta\alpha}\alpha_y + u_{\beta\beta}\beta_y = u_{\alpha\beta} - u_{\alpha\alpha}.$$

Подставим u_{xx}, u_{xy}, u_{yy} в заданное уравнение

$$u_{\alpha\alpha} - 2u_{\alpha\beta} + u_{\beta\beta} + 2u_{\alpha\beta} - 2u_{\alpha\alpha} + 2u_{\alpha\alpha} = 0.$$

Окончательно получим

$$u_{\alpha\alpha} + u_{\beta\beta} = 0.$$

§ 14. РЕШЕНИЕ ОДНОМЕРНОГО ВОЛНОВОГО УРАВНЕНИЯ

Как уже говорилось, при решении задач математической физики часто встречается волновое уравнение. Одномерное волновое уравнение может быть однородным

$$u_{tt} - a^2 u_{xx} = 0 \tag{14}$$

или неоднородным

$$u_{tt} - a^2 u_{xx} = f(x, t). \tag{15}$$

Здесь t — время, x — линейная координата, a — постоянная величина, $f(x, t)$ — известная функция от x и t.

В частности, уравнение (9) для малых поперечных колебаний струны является однородным. Для коротких высокочастотных воздушных линий и коаксиальных кабелей, применяемых в радиотехнике, где с достаточной степенью точности можно пренебречь сопро-

тивлением r_0 и проводимостью g_0 ($r_0 \approx 0$ и $g_0 \approx 0$), для линий без потерь, из уравнений (5) и (6) получим также однородное волновое уравнение вида

$$u_{tt} - \frac{1}{L_0 C_0} u_{xx} = 0, \tag{16}$$

где

$$a^2 = \frac{1}{L_0 C_0}. \tag{17}$$

Существует несколько различных способов решения волнового уравнения. Одним из наиболее распространенных методов решения является метод Фурье (метод разделения переменных).

Покажем, как методом Фурье решается неоднородное волновое уравнение (15)

$$u_{tt} - a^2 u_{xx} = f(x, t)$$

при начальных условиях (10) и (11)

$$u(x, 0) = \varphi(x), \quad u_t(x, 0) = \psi(x)$$

и граничных условиях (12) и (13)

$$u(0, t) = 0, \quad u(l, t) = 0.$$

Будем искать решение в виде

$$u(x, t) = \sum_{n=1}^{\infty} T_n(t) \sin \frac{n\pi x}{l}. \tag{18}$$

Ясно, что функция $u(x, t)$ при любой функции $T_n(t)$ удовлетворяет граничным условиям, так как она равна нулю при $x = 0$ и при $x = l$. Выясним, каким условиям должны удовлетворять функции $T_n(t)$ для того, чтобы функция $u(x, t)$ удовлетворяла заданным начальным условиям.

Используя известные начальные условия, получаем:

$$\varphi(x) = \sum_{n=1}^{\infty} T_n(0) \sin \frac{n\pi x}{l},$$

$$\psi(x) = \sum_{n=1}^{\infty} T_n'(0) \sin \frac{n\pi x}{l}.$$

Эти равенства представляют собой разложение в ряд Фурье по синусам функций $\varphi(x)$ и $\psi(x)$.

Тогда

$$T_n(0) = \frac{2}{l} \int_0^l \varphi(x) \sin \frac{n\pi x}{l} dx,$$

$$T_n'(0) = \frac{2}{l} \int_0^l \psi(x) \sin \frac{n\pi x}{l} dx. \tag{19}$$

Потребуем, чтобы функция $u(x, t)$ (18) удовлетворяла уравнению (15). Для подстановки в указанное уравнение соответствующих величин найдем

$$u_{tt}(x, t) = \sum_{n=1}^{\infty} T_n''(t) \sin \frac{n\pi x}{l} \tag{20}$$

и

$$u_{xx}(x, t) = -\sum_{n=1}^{\infty} T_n(t) \left(\frac{n\pi}{l}\right)^2 \sin \frac{n\pi x}{l}. \tag{21}$$

Функцию $f(x, t)$ разложим в ряд Фурье по синусам:

$$f(x, t) = \sum_{n=1}^{\infty} b_n(t) \sin \frac{n\pi x}{l}, \tag{22}$$

где

$$b_n(t) = \frac{2}{l} \int_0^l f(x, t) \sin \frac{n\pi x}{l} dx \tag{23}$$

(при этом аргумент t рассматривается как параметр).

Подставив значения u_{tt}, u_{xx}, $f(x, t)$ из формул (20), (21) и (22) в уравнение (15), получим

$$\sum_{n=1}^{\infty} \left(T_n''(t) + \left(\frac{an\pi}{l}\right)^2 T_n(t) - b_n(t)\right) \sin \frac{n\pi x}{l} = 0.$$

Следовательно, для того чтобы функция $u(x, t)$ (18) удовлетворяла уравнению (15) и начальным условиям (10), (11), функции $T_n(t)$ должны удовлетворять дифференциальным уравнениям

$$T_n''(t) + \left(\frac{an\pi}{l}\right)^2 T_n(t) = b_n(t) \tag{24}$$

и своим начальным условиям (19).

Таким образом, нахождение решения уравнения (15), удовлетворяющего начальным условиям (10) и (11), а также граничным условиям (12) и (13), сводится к решению обыкновенных дифференциальных уравнений (24). Последние являются линейными дифференциальными уравнениями второго порядка с постоянными коэффициентами.

Пример. Найти решение уравнения $u_{tt} - u_{xx} = 1$, удовлетворяющее начальным условиям $u(x, 0) = 0$, $u_t(x, 0) = 1$ и граничным условиям $u(0, t) = 0$, $u(1, t) = 0$.

Решение. Записываем решение в виде:

$$u(x, t) = \sum_{n=1}^{\infty} T_n(t) \sin n\pi x \quad (l = 1).$$

Правую часть (неоднородность) $f(x, t) \equiv 1$ раскладываем в ряд Фурье по синусам, для чего находим коэффициенты этого ряда:

$$b_n = 2\int_0^1 f(x, t)\sin n\pi x dx = 2\int_0^1 \sin n\pi x dx = -\frac{2}{n\pi}\cos n\pi x\Big|_0^1 = \frac{2}{n\pi}(1 - \cos n\pi);$$

$$b_n = \begin{cases} 0, & \text{если } n = 2k, \\ \dfrac{4}{n\pi}, & \text{если } n = 2k-1. \end{cases} \quad k \in N,$$

Для функций $T_n(t)$ получим дифференциальные уравнения

$$T_n''(t) + (n\pi)^2 T_n(t) = b_n \quad (a = 1,\ l = 1).$$

Чтобы записать начальные условия для этих дифференциальных уравнений, надо найти коэффициенты разложения функций $u(x, 0)$ и $u_t(x, 0)$ в ряды Фурье по синусам.

В данном примере $u(x, 0) \equiv 0$, поэтому и коэффициенты ряда Фурье этой функции равны нулю. Следовательно, $T_n(0) = 0$.

Так как $u_t(x, 0) \equiv 1$, то из равенства (19) имеем

$$T_n'(0) = 2\int_0^1 u_t(x, 0)\sin n\pi x dx = 2\int_0^1 \sin n\pi x dx = \begin{cases} 0, & \text{если } n = 2k, \\ \dfrac{4}{n\pi}, & \text{если } n = 2k-1. \end{cases}$$

Пусть $n = 2k$. Тогда для нахождения $T_{2k}(t)$ надо решить однородное линейное дифференциальное уравнение

$$T_{2k}'' + (2k\pi)^2 T_{2k} = 0$$

при начальных условиях $T_{2k}(0) = 0$, $T_{2k}'(0) = 0$.

Легко проверить, что $T_{2k}(t) \equiv 0$.

Пусть $n = 2k - 1$. В этом случае надо решить уравнение

$$T_{2k-1}'' + ((2k-1)\pi)^2 T_{2k-1} = \frac{4}{(2k-1)\pi}$$

при начальных условиях

$$T_{2k-1}(0) = 0, \quad T_{2k-1}'(0) = \frac{4}{(2k-1)\pi}.$$

Общее решение этого уравнения имеет вид

$$T_{2k-1}(t) = C_1\cos(2k-1)\pi t + C_2\sin(2k-1)\pi t + \frac{4}{(2k-1)^3\pi^3}.$$

Положив $t = 0$ и учитывая, что $T_{2k-1}(0) = 0$, получим

$$0 = C_1 + \frac{4}{(2k-1)^3\pi^3}, \quad C_1 = -\frac{4}{(2k-1)^3\pi^3}.$$

Имеем

$$T_{2k-1}'(t) = -C_1(2k-1)\pi\sin(2k-1)\pi t + C_2(2k-1)\pi\cos(2k-1)\pi t.$$

Положив здесь $t = 0$ и учитывая что $T_{2k-1}'(0) = \dfrac{4}{(2k-1)\pi}$, получим

$$\frac{4}{(2k-1)\pi} = C_2(2k-1)\pi,$$

отсюда

$$C_2 = \frac{4}{(2k-1)^2 \pi^2},$$

тогда

$$T_{2k-1}(t) = \frac{4}{(2k-1)^3 \pi^3} ((2k-1)\pi \sin(2k-1)\pi t - \cos(2k-1)\pi t + 1).$$

Следовательно,

$$u(x, t) = \frac{4}{\pi^3} \sum_{k=1}^{\infty} \frac{(2k-1)\pi \sin(2k-1)\pi t - \cos(2k-1)\pi t + 1}{(2k-1)^3} \sin(2k-1)\pi x.$$

При установившемся синусоидальном режиме в однородной двухпроводной линии целесообразно осуществить разделение переменных в уравнениях (3) и (4) с помощью изображения мгновенных значений тока i и напряжения u (на расстоянии x от начала линии) комплексными числами (в электротехнике символ i ($i^2 = -1$) обозначают буквой j)

$$i(x, t) \doteqdot \dot{I}(x) e^{j\omega t} = \dot{I} e^{j\omega t},$$

$$u(x, t) \doteqdot \dot{U}(x) e^{j\omega t} = \dot{U} e^{j\omega t},$$

где $\dot{I} = \dot{I}(x) = I e^{j\psi_i}$ — комплекс действующего значения тока; $\dot{U} = \dot{U}(x) = U e^{j\psi_u}$ — комплекс действующего значения напряжения; ω — угловая частота периодических (строго синусоидальных во времени t) тока и напряжения, с действующими значениями $I = \frac{I_m}{\sqrt{2}}$ и $U = \frac{U_m}{\sqrt{2}}$ и с начальными фазами (при $t = 0$) соответственно ψ_i и ψ_u.

Так как $\dot{I}$ и $\dot{U}$ не зависят от времени, а являются только функциями x, то уравнения (3) и (4) однородной линии с частными производными

$$\begin{cases} -\frac{\partial i}{\partial x} = g_0 u + C_0 \frac{\partial u}{\partial t}, \\ -\frac{\partial u}{\partial x} = r_0 i + L_0 \frac{\partial i}{\partial t} \end{cases}$$

после подстановки значений, изображающих $i(x, t)$ и $u(x, t)$, и дифференцирования преобразуются в обыкновенные дифференциальные уравнения

$$-\frac{d\dot{I}}{dx} e^{j\omega t} = g_0 \dot{U} e^{j\omega t} + j\omega C_0 \dot{U} e^{j\omega t}$$

$$-\frac{d\dot{U}}{dx} e^{j\omega t} = r_0 \dot{I} e^{j\omega t} + j\omega L_0 \dot{I} e^{j\omega t},$$

которые после сокращения на множитель $e^{j\omega t} \neq 0$ дают систему уравнений:

$$\begin{cases} -\frac{d\dot{I}}{dx} = (g_0 + j\omega C_0)\dot{U}, & (25) \\ -\frac{d\dot{U}}{\partial x} = (r_0 + j\omega L_0)\dot{I}. & (26) \end{cases}$$

Эту систему уравнений (25) и (26) можно свести к уравнению, содержащему только искомый ток $\dot{I}(x)$, или к уравнению, содержащему только искомое напряжение $\dot{U}(x)$.

Действительно, если продифференцировать все члены уравнения (25) по переменной x и в результат подставить значение $\frac{d\dot{U}}{dx}$ из уравнения (26), то получим дифференциальное уравнение второго порядка относительно $\dot{I}$

$$\frac{d^2\dot{I}}{\partial x^2} = (r_0 + j\omega L_0)(g_0 + i\omega C_0)\dot{I},$$

или

$$\frac{d^2\dot{I}}{dx^2} - \gamma^2\dot{I} = 0, \quad (27)$$

где

$$\gamma = \sqrt{(r_0 + j\omega L_0)(g_0 + j\omega C_0)} = \alpha + j\beta \quad (28)$$

называется коэффициентом распространения.

Аналогично, исключая $\dot{I}$ из уравнения (26), получим уравнение относительно $\dot{U}$

$$\frac{d^2\dot{U}}{dx^2} = (r_0 + j\omega L_0)(g_0 + j\omega C_0)\dot{U},$$

или

$$\frac{d^2\dot{U}}{dx^2} - \gamma^2\dot{U} = 0. \quad (29)$$

Дифференциальные уравнения (27) и (29), определяющие изменение комплексов действующих значений тока и напряжения вдоль линии, одинаковы. Поэтому достаточно определить, например, закон изменения напряжения $\dot{U}$. Тогда закон изменения тока $\dot{I}$ можно получить с помощью уравнения (26).

Известно, что решение однородного линейного дифференциального уравнения второго порядка с постоянными коэффициентами типа (27) или (29) имеет вид

$$\dot{U} = A_1 e^{-\gamma x} + A_2 e^{\gamma x}, \quad (30)$$

где A_1 и A_2 — комплексные постоянные интегрирования, определяемые из граничных условий.

При этом ток $\dot{I}$ определится из уравнения (26) следующим образом:

$$\dot{I} = -\frac{1}{r_0 + j\omega L_0} \cdot \frac{d\dot{U}}{dx} =$$

$$= \frac{\gamma}{r_0 + j\omega L_0} (A_1 e^{-\gamma x} - A_2 e^{\gamma x}) = \frac{1}{\sqrt{\frac{r_0 + j\omega L_0}{g_0 + j\omega C_0}}} (A_1 e^{-\gamma x} - A_2 e^{\gamma x}),$$

или

$$\dot{I} = \frac{1}{Z_C} (A_1 e^{-\gamma x} - A_2 e^{\gamma x}), \tag{31}$$

где $Z_C = \sqrt{\frac{r_0 + j\omega L_0}{g_0 + j\omega C_0}}$ — комплексная величина, имеющая размерность сопротивления и называемая волновым или характеристическим сопротивлением однородной линии.

В показательной форме

$$Z_C = \sqrt{\frac{r_0 + j\omega L_0}{g_0 + j\omega C_0}} = z_C e^{j\theta}. \tag{32}$$

Запишем постоянные интегрирования A_1 и A_2 в показательной форме:

$$A_1 = |A_1| e^{j\psi_1}, \quad A_2 = |A_2| e^{j\psi_2}$$

и подставим их значения, а также значения γ из формулы (28) и Z_C из равенства (32) в выражения (30) и (31). Получим:

$$\dot{U} = |A_1| e^{-\alpha x} \cdot e^{j(-\beta x + \psi_1)} + |A_2| e^{\alpha x} \cdot e^{j(\beta x + \psi_2)}, \tag{33}$$

$$\dot{I} = \frac{|A_1|}{z_C} e^{-\alpha x} \cdot e^{j(-\beta x + \psi_1 - \theta)} - \frac{|A_2|}{z_C} e^{\alpha x} \cdot e^{j(\beta x + \psi_2 - \theta)}. \tag{34}$$

Мгновенные значения напряжения $u(x, t)$ и тока $i(x, t)$ в текущей точке, отстоящей на расстоянии x от начала линии, равны мнимой части выражений $\sqrt{2}\dot{U}e^{j\omega t}$ и $\sqrt{2}\dot{I}e^{j\omega t}$. Поэтому искомые величины $u(x, t)$ и $i(x, t)$ с учетом зависимостей (33) и (34) запишутся так:

$$u(x, t) = \sqrt{2}|A_1| e^{-\alpha x} \sin(\omega t - \beta x + \psi_1) + \\ + \sqrt{2}|A_2| e^{\alpha x} \sin(\omega t + \beta x + \psi_2) \tag{35}$$

и

$$i(x, t) = \frac{\sqrt{2}|A_1|}{z_C} e^{-\alpha x} \sin(\omega t - \beta x + \psi_1 - \theta) - \\ - \frac{\sqrt{2}|A_2|}{z_C} e^{\alpha x} \sin(\omega t + \beta x + \psi_2 - \theta). \tag{36}$$

Выражение (35) представляет собой решение неоднородного волнового уравнения длинной линии с потерями.

Из уравнений (35) и (36) следует, что мгновенные значения напряжения и тока в любом сечении состоят из двух слагаемых, каждое из которых является функцией двух переменных x и t.

Рассмотрим последовательно каждое из слагаемых $u(x, t)$, входящих в уравнение (35). Обозначим первое из них u_n, а второе u_0:

$$u_n = \sqrt{2}\,|A_1|\,e^{-\alpha x}\sin(\omega t - \beta x + \psi_1), \tag{37}$$

$$u_0 = \sqrt{2}\,|A_2|\,e^{\alpha x}\sin(\omega t + \beta x + \psi_2). \tag{38}$$

Из выражения для u_n (37) видно, что если точку x взять фиксированной ($x = x_1$), то эта составляющая напряжения будет изменяться во времени в данной точке по строго синусоидальному закону с постоянной амплитудой $\sqrt{2}\,|A_1|\,e^{-\alpha x_1}$ и частотой ω.

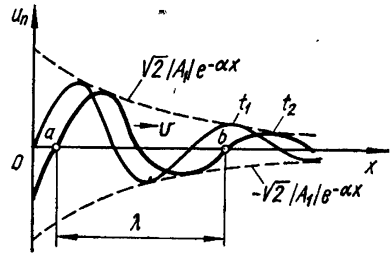

Рис. 35

Если же принять момент времени t фиксированным ($t = t_1$), то изменение u_n вдоль линии в заданный момент времени в зависимости от x будет происходить по закону затухающей синусоиды с амплитудой $\sqrt{2}\,|A_1|\,e^{-\alpha x}$, убывающей по показательному закону с ростом x, т. е. по мере удаления текущей точки от начала к концу линии. При новом фиксированном значении $t_2 > t_1$ характер изменения $u_n(x)$ останется прежним, только затухающая синусоида в этом случае переместится по отношению к затухающей синусоиде при $t = t_1$ вдоль оси Ox на некоторое расстояние (рис. 35).

Таким образом, составляющая напряжения $u_n(x, t)$ представляет собой бегущую волну, движущуюся от начала к концу линии и затухающую в направлении ее движения. Эту волну называют прямой, или падающей волной.

Длина бегущей волны λ определяется расстоянием между двумя ближайшими точками, фазы колебаний которых отличаются на 2π. Из выражения (37) следует, что для отмеченных на рис. 35 точек a и b имеем

$$(\omega t - \beta x + \psi_1) - (\omega t - \beta(x + \lambda) + \psi_1) = 2\pi,$$

откуда

$$\lambda = \frac{2\pi}{\beta}, \tag{39}$$

где коэффициент фазы β зависит от частоты и параметров линии.

Нетрудно показать (из выражения для γ), что коэффициент β определяется по формуле

$$\beta = \sqrt{\frac{1}{2}(\omega^2 L_0 C_0 - r_0 g_0 + \sqrt{(r_0^2 + \omega^2 L_0^2)(g_0^2 + \omega^2 C_0^2)})}. \tag{40}$$

Скорость движения бегущей волны называется фазовой скоростью. Она определяется как скорость v перемещения точки, фаза колебания в которой остается неизменной. Следовательно, для прямой волны u_n имеем

$$\omega t - \beta x + \psi_1 = \text{const}.$$

Тогда

$$\frac{d}{dt}(\omega t - \beta x + \psi_1) = 0$$

и

$$v = \frac{dx}{dt} = \frac{\omega}{\beta}. \tag{41}$$

Из выражений для λ и v бегущей волны видно, что в общем случае они зависят как от параметров длинной линии, так и от частоты ω колебаний напряжения и тока.

Рассматривая аналогично второе слагаемое $u(x, t)$, входящее в уравнение (35) и записанное формулой (38) для $u_0(x, t)$, делаем вывод, что составляющая напряжения $u_0(x, t)$ представляет собой также бегущую волну, но движущуюся от конца к началу линии и затухающую в этом же направлении. Эту волну называют обратной, или отраженной волной. Фазовая скорость обратной волны равна по величине фазовой скорости прямой волны, но противоположна ей по знаку. Длины обеих волн u_n и u_0 одинаковы. Одинакова также и интенсивность затухания этих волн в направлении их движения, которая характеризуется величиной α, называемой коэффициентом затухания.

Из выражения для γ (28) можно показать, что коэффициент затухания α определяется по формуле

$$\alpha = \sqrt{\frac{1}{2}(r_0 g_0 - \omega^2 L_0 C_0 + \sqrt{(r_0^2 + \omega^2 L_0^2)(g_0^2 + \omega^2 C_0^2)})}. \tag{42}$$

Следовательно, этот коэффициент, как и β, зависит от первичных параметров линии и от частоты.

Вышеизложенное позволяет заключить, что мгновенное значение напряжения $u(x, t)$ в линии можно рассматривать как результат наложения прямой и обратной бегущих волн, движущихся навстречу друг другу и затухающих в направлении их движения. То же самое можно сказать и о мгновенном значении тока в линии $i(x, t)$, только прямая и обратная волны тока не складываются, как для напряжения, а вычитаются. Однако в обоих случаях при наложении бегущих волн результирующие кривые распределения мгновенных значений напряжений и токов вдоль линии в любой фиксированный момент времени имеют волнообразный характер.

Ясно, что физически в линии существуют только результирующие ток $i(x, t)$ и напряжение $u(x, t)$. Разложение же их на прямые и обратные волны является лишь удобным приемом, облегчающим исследование установившегося синусоидального режима в однородных длинных линиях.

Рассмотрим еще один метод решения волнового уравнения — метод д'Аламбера.

Возьмем случай, когда граничные условия нас не интересуют или когда их можно не учитывать (например, в случае неограниченной струны, бесконечно длинной линии, неограниченного стержня и т. д.). В этих случаях задача ставится так:

Найти решение волнового уравнения

$$u_{tt} - a^2 u_{xx} = 0 \quad (t = y,\ a_{11} = -a^2,\ a_{12} = 0,\ a_{22} = 1),$$

удовлетворяющее начальным условиям

$$u(x, 0) = \varphi(x); \quad u_t(x, 0) = \psi(x),$$

где $\varphi(x)$ и $\psi(x)$ — заданные в $]-\infty; +\infty[$ функции.

Приведем волновое уравнение к каноническому виду, содержащему смешанную производную. Здесь характеристическое уравнение

$$a_{11}dt^2 - 2a_{12}dxdt + a_{22}dx^2 = 0$$

примет вид

$$-a^2dt^2 + dx^2 = 0,$$

или

$$dx^2 - a^2dt^2 = 0.$$

Оно распадается на два уравнения:

$$dx - adt = 0 \quad \text{и} \quad dx + adt = 0,$$

интегралы которых будут $x - at = C_1$, $x + at = C_2$.

Введем новые переменные:

$$\xi = x - at, \quad \eta = x + at.$$

Тогда

$$\xi_x = 1, \quad \xi_t = -a, \quad \eta_x = 1, \quad \eta_t = a,$$

$$u_x = u_\xi \xi_x + u_\eta \eta_x = u_\xi + u_\eta,$$

$$u_{xx} = u_{\xi\xi}\xi_x + u_{\xi\eta}\eta_x + u_{\eta\xi}\xi_x + u_{\eta\eta}\eta_x = u_{\xi\xi} + 2u_{\xi\eta} + u_{\eta\eta},$$

$$u_t = u_\xi \xi_t + u_\eta \eta_t = -au_\xi + au_\eta,$$

$$u_{tt} = -au_{\xi\xi}\xi_t - au_{\xi\eta}\eta_t + au_{\eta\xi}\xi_t + au_{\eta\eta}\eta_t = a^2u_{\xi\xi} - 2a^2u_{\xi\eta} + a^2u_{\eta\eta}.$$

Подставив u_{xx}, u_{tt} в исходное уравнение, получим

$$a^2u_{\xi\xi} - 2a^2u_{\xi\eta} + a^2u_{\eta\eta} - a^2(u_{\xi\xi} + 2u_{\xi\eta} + u_{\eta\eta}) = 0,$$

$$-4a^2u_{\xi\eta} = 0,$$

$$u_{\xi\eta} = 0.$$

Полученное уравнение можно записать так:

$$\frac{\partial}{\partial\xi}(u_\eta) = 0.$$

Отсюда следует, что u_η не зависит от ξ:

$$u_\eta = f^*(\eta),$$

где $f^*(\eta)$ — произвольная функция η.

Интегрируя последнее равенство по η при фиксированном ξ, имеем

$$u(\xi, \eta) = \int f^*(\eta)\, d\eta + f_1(\xi) = f_1(\xi) + f_2(\eta),$$

где $f_1(\xi)$ и $f_2(\eta)$ — произвольные дважды дифференцируемые функции аргументов ξ и η.

Учитывая, что $\xi = x - at$ и $\eta = x + at$, получаем общее решение данного уравнения в виде

$$u(x, t) = f_1(x - at) + f_2(x + at).$$

Определим функции f_1 и f_2 так, чтобы функция $u(x, t)$ удовлетворяла начальным условиям:

$$u(x, 0) = f_1(x) + f_2(x) = \varphi(x),$$

$$u_t(x, 0) = -af_1'(x) + af_2'(x) = \psi(x).$$

Таким образом, для нахождения функций f_1 и f_2 имеем систему уравнений

$$\begin{cases} f_1(x) + f_2(x) = \varphi(x), \\ -af_1'(x) + af_2'(x) = \psi(x). \end{cases}$$

Интегрируя второе равенство, получаем

$$-f_1(x) + f_2(x) = \frac{1}{a}\int_{x_0}^{x} \psi(z)\, dz + C,$$

где x_0 и C — постоянные. Тогда

$$\begin{cases} f_1(x) + f_2(x) = \varphi(x), \\ -f_1(x) + f_2(x) = \frac{1}{a}\int_{x_0}^{x} \psi(z)\, dz + C. \end{cases}$$

Отсюда находим

$$f_1(x) = \frac{1}{2}\varphi(x) - \frac{1}{2a}\int_{x_0}^{x} \psi(z)\, dz - \frac{C}{2}$$

и

$$f_2(x) = \frac{1}{2}\varphi(x) + \frac{1}{2a}\int_{x_0}^{x} \psi(z)\, dz + \frac{C}{2}.$$

Подставив в выражение для $u(x, t)$ найденные значения f_1 и f_2, получим

$$u(x, t) = \frac{\varphi(x - at) + \varphi(x + at)}{2} + \frac{1}{2a}\int_{x_0}^{x+at} \psi(z)\, dz - \frac{1}{2a}\int_{x_0}^{x-at} \psi(z)\, dz,$$

или

$$u(x, t) = \frac{\varphi(x - at) + \varphi(x + at)}{2} + \frac{1}{2a}\int_{x-at}^{x+at} \psi(z)\, dz. \qquad (43)$$

Это равенство называется формулой д'Аламбера.

Для того чтобы истолковать формулу, дающую решение волнового уравнения методом д'Аламбера, рассмотрим предварительно понятие волны.

Пусть в пространстве определено переменное поле $u(x, y, z, t)$, где u — скалярная функция. Это поле $u(x, y, z, t)$ называется волной, если заданную функцию можно представить в виде $u = u(v, t)$, где $v = v(x, y, z)$ зависит только от координат. Величину v называют пространственным аргументом функции u, а саму функцию $u = u(v, t)$ — волновой функцией. Поверхность $v(x, y, z) = \text{const}$ называется фазовой поверхностью или фронтом волны. Фазовая поверхность есть геометрическое место точек, в которых пространственный аргумент v волновой функции $u = u(v, t)$ имеет одно и то же значение. В зависимости от вида фронта волны различают плоские, цилиндрические и сферические волны.

Пример. Задано переменное поле $u = x^2 + y^2 + z^2 - a^2t$. Здесь функция $u(x, y, z, t)$ волновая, так как можно записать $u = v - a^2t$, где $v = x^2 + y^2 + z^2$.

В данном случае фазовые поверхности — концентрические сферы с центром в начале координат. Следовательно, нам задана сферическая волна.

В одномерном случае, когда пространственный аргумент v зависит только от одной из координат (например, от x), функция $u = u(x, t)$ всегда является плоской волной.

Если волновую функцию можно записать в виде

$$u(v, t) = V(x, y, z) \cdot \psi(t),$$

где $V(x, y, z)$ — пространственный множитель, а $\psi(t)$ — временной множитель, то волна называется стоячей.

Характерной особенностью стоячей волны является то, что закон колебания в этом случае определяется временным множителем $\psi(t)$. Пространственный же множитель $\varphi(v) = V(x, y, z)$ определяет лишь масштаб колебаний.

Фазовые поверхности, на которых пространственный множитель равен нулю, называются узловыми поверхностями (узлами, если пространственный множитель зависит только от одной координаты).

Пример. Известно, что решение одномерного однородного уравнения, найденное методом Фурье, имеет вид

$$u(x, t) = \sum_{n=1}^{\infty} \left(A_n \cos\frac{a_n \pi t}{l} + B_n \sin\frac{a_n \pi t}{l}\right) \sin\frac{n\pi x}{l}.$$

Определить, что представляет собой эта волновая функция.

Р е ш е н и е. Здесь каждый член ряда

$$\left(A_n \cos \frac{a_n \pi t}{l} + B_n \sin \frac{a_n \pi t}{l}\right) \sin \frac{n \pi x}{l}, \quad n \in N,$$

представляет собой стоячую волну:

$$V = \sin \frac{n \pi x}{l}, \quad \psi = A_n \cos \frac{a_n \pi t}{l} + B_n \sin \frac{a_n \pi t}{l}.$$

Узлами этой волны являются точки x, где $\sin \frac{n \pi x}{l} = 0$, т. е. точки

$$x = \frac{ml}{n} \quad (m = 1, 2, \ldots, n - 1).$$

Таким образом, $u(x, t)$ представляет собой бесконечную сумму стоячих волн.

Если волновая функция u может быть записана в виде

$$u = u(v - at),$$

где a — постоянная, то волна называется б е г у щ е й в о л н о й б е з и с к а ж е н и й. Пусть $v_1(x, y, z) = v_1$, $v_2(x, y, z) = v_2$ — две фазовые поверхности. Причем колебания функции u на этих поверхностях с течением времени t определяются уравнениями

$$u = u(v_1 - at) \quad \text{и} \quad u = u(v_2 - at).$$

Тогда, записав второе из этих уравнений в виде

$$u = u\left(v_1 - a\left(t - \frac{v_2 - v_1}{a}\right)\right),$$

приходим к выводу, что колебания функции u на второй фазовой поверхности $v = v_2$ будут в точности повторять колебания на первой поверхности $v = v_1$, но с запаздыванием на промежуток времени $\frac{v_2 - v_1}{a}$. С физической точки зрения это означает, что здесь изменяется только начальная фаза колебаний.

Чтобы представить графически процесс распространения возмущения u, изобразим на плоскости vOu функцию u в начальный момент времени $t = 0$ в виде $u = u(v)$ и в какой-нибудь другой фиксированный момент времени t в виде $u = u(v - at)$. Пусть $v = at + \tilde{v}$, т. е. перенесем начало координат O вдоль оси v в новую точку $O_1(at, 0)$. Тогда получим

$$u = u(\tilde{v}).$$

Отсюда вытекает, что графически функция u в начальный момент времени и в момент времени t имеет один и тот же вид, только вторая кривая смещена относительно первой вдоль оси Ov на расстояние at (рис. 36).

График функции $u = u(v)$, определяющий распределение функции u в начальный момент времени, называют п р о ф и л е м в о л-

ны. С течением времени этот график на плоскости vOu, не изменяя своей формы, перемещается вдоль оси Ov с постоянной скоростью a.

В частности, бегущая волна без искажений будет иметь место при установившемся синусоидальном режиме в линии без потерь, замкнутой на характеристическое сопротивление $Z_2 = \sqrt{\frac{L_0}{C_0}} = z_C$. При этом профиль волны будет синусоидальным.

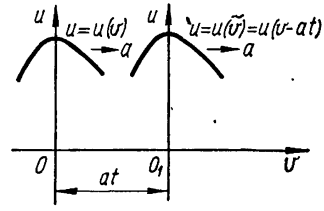

Рис. 36

Вернемся теперь к формуле д'Аламбера (43). Она дает решение одномерного однородного волнового уравнения и имеет вид

$$u(x, t) = \frac{\varphi(x - at) + \varphi(x + at)}{2} + \frac{1}{2a}\int_{x-at}^{x+at} \psi(z)\,dz.$$

Ранее функцию $u(x, t)$ мы записывали так:

$$u(x, t) = f_1(x - at) + f_2(x + at),$$

где первое слагаемое

$$f_1(x - at) = \frac{1}{2}\varphi(x - at) - \frac{1}{2a}\int_{x_0}^{x-at} \psi(z)\,dz$$

при $x - at = \text{const}$ сохраняет постоянное значение (оно остается неизменным, когда наблюдатель движется вдоль оси Ox в положительном направлении со скоростью a). Следовательно, функция $f_1(x - at)$ описывает распространение прямой бегущей волны без искажений.

Аналогично функция

$$f_2(x + at) = \frac{1}{2}\varphi(x + at) + \frac{1}{2a}\int_{x_0}^{x+at} \psi(z)\,dz$$

представляет обратную бегущую волну без искажений, распространяющуюся с той же постоянной скоростью a, но в отрицательном направлении оси Ox.

В целом процесс распространения колебаний, описываемый функцией $u(x, t)$, представляет собой с у п е р п о з и ц и ю (наложение) прямой и обратной бегущих волн без искажений.

§ 15. НЕКОТОРЫЕ СПЕЦИАЛЬНЫЕ ФУНКЦИИ

В математике и ее приложениях часто встречаются функции, не выражающиеся через известные элементарные функции. Их называют с п е ц и а л ь н ы м и функциями. Они достаточно освещены в математической литературе, здесь же мы рассмотрим лишь некоторые из этих функций.

Функция гамма (Г-функция). Запишем интеграл

$$\int_0^{+\infty} e^{-x}x^{\nu-1}dx.$$

Это интеграл Эйлера второго рода. Можно показать, что при $\nu > 0$ он сходится. Тогда этот интеграл представляет собой функцию от ν. Ее обозначают $\Gamma(\nu)$ и называют функцией гамма. Таким образом,

$$\Gamma(\nu) = \int_0^{+\infty} e^{-x}t^{\nu-1}dx, \tag{44}$$

где $\nu > 0$.

При $\nu \leqslant 0$ интеграл Эйлера расходится, и функция гамма с помощью этого интеграла (при таких ν) не может быть определена.

Положив в интеграле Эйлера $x = t^2$, получим

$$\Gamma(\nu) = 2\int_0^{+\infty} e^{-t^2}t^{2\nu-1}dt. \tag{45}$$

По формуле (44) при $\nu = 1$ имеем

$$\Gamma(1) = \int_0^{+\infty} e^{-x}dx = -e^{-x}\Big|_0^{+\infty} = 1.$$

Следовательно, $\Gamma(1) = 1$. Приняв в формуле (45) $\nu = \frac{1}{2}$, найдем

$$\Gamma\left(\frac{1}{2}\right) = 2\int_0^{+\infty} e^{-t^2}dt = 2\frac{\sqrt{\pi}}{2} = \sqrt{\pi}$$

$\left(\text{известно, что } \int_0^{+\infty} e^{-t^2}dt = \frac{1}{2}\sqrt{\pi}\right)$.

К интегралу (44) применим формулу интегрирования по частям, считая $u = x^{\nu-1}$, $dv = e^{-x}\,dx$. Тогда после вычислений получим, что

$$\Gamma(\nu) = (\nu - 1)\cdot\Gamma(\nu - 1) \quad (\nu > 1). \tag{46}$$

Это основная формула приведения для функции гамма. Из нее следует, что

$$\Gamma(\nu - 1) = \frac{\Gamma(\nu)}{\nu - 1}.$$

Найдем

$$\lim_{\substack{\nu\to 1\\ \nu>1}} \Gamma(\nu - 1) = \lim_{\substack{\nu\to 1\\ \nu>1}} \frac{\Gamma(\nu)}{\nu - 1} = \lim_{\substack{\nu\to 1\\ \nu>1}} \frac{1}{\nu - 1} = +\infty,$$

так как $\lim_{\nu \to 1} \Gamma(\nu) = \Gamma(1) = 1$. Применив формулу (46) последовательно k раз, получим

$$\Gamma(\nu) = (\nu - 1)(\nu - 2) \ldots (\nu - k)\,\Gamma(\nu - k) \quad (\nu - k > 0). \tag{47}$$

В математических справочниках значения функции гамма даются обычно лишь для величин ν, лежащих в [1; 2].

Чтобы найти значение функции $\Gamma(\nu)$ при $0 < \nu < 1$, надо применить формулу (47). Для нахождения $\Gamma(\nu)$ при $\nu > 2$ мы также применяем формулу (47), выбрав целое $k > 0$, так, чтобы $1 \leqslant \nu - k < 2$. Если $\nu = n$, где $n > 0$ — целое, то

$$\Gamma(n) = (n - 1)!.$$

Применяя формулу (47) для $\nu = n + \frac{1}{2}$ и учитывая, что $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, получаем

$$\Gamma\left(n + \frac{1}{2}\right) = \frac{(2n-1)!!}{2^n}\sqrt{\pi},$$

где $(2n - 1)!! = 1 \cdot 3 \cdot 5 \ldots (2n - 1)$. До сих пор мы считали, что аргумент ν функции $\Gamma(\nu)$ больше нуля. Доопределим теперь функцию гамма для отрицательных значений аргумента. Из формулы (46) найдем

$$\Gamma(\nu - k) = \frac{\Gamma(\nu)}{(\nu - k)(\nu - k + 1) \ldots (\nu - 2)(\nu - 1)}.$$

Положим $\nu - k = \nu^*$, тогда

$$\Gamma(\nu^*) = \frac{\Gamma(\nu^* + k)}{\nu^*(\nu^* + 1) \ldots (\nu^* + k - 1)}.$$

Обозначив в последней формуле ν^* снова через ν, получим

$$\Gamma(\nu) = \frac{\Gamma(\nu + k)}{\nu(\nu + 1) \ldots (\nu + k - 1)}. \tag{48}$$

Если $\nu + k > 0$ и $\nu \neq -1, -2, -3$ и т. д., $k \in N$, то правая часть формулы имеет смысл и при $\nu < 0$.

Формулу (48) принимают за определение функции гамма при отрицательных значениях аргумента ν. Очевидно, функция гамма не существует при целых отрицательных значениях ν (при таких значениях ν она обращается в бесконечность). Чтобы найти значения функции гамма для отрицательных значений ее аргумента ν, пользуясь таблицей, нужно взять $k > 0$ целое таким, чтобы $-\nu < k < -\nu + 1$, и применить формулу (48).

Цилиндрические функции. *Цилиндрическими функциями называются решения уравнения Бесселя*

$$x^2y'' + xy' + (x^2 - \nu^2)y = 0.$$

Если искать решение этого уравнения в виде обобщенного степенного ряда

$$y = \sum_{k=0}^{+\infty} a_k x^{\rho+k},$$

то после подстановки этого ряда в уравнение Бесселя и определения ρ и a_k, получим

$$y = a_0 \sum_{k=0}^{+\infty} (-1)^k \frac{x^{\nu+2k}}{2^{2k} \cdot k! \, (\nu+1)(\nu+2) \ldots (\nu+k)}. \qquad (49)$$

Полагая здесь

$$a_0 = \frac{1}{2^\nu \cdot \Gamma(\nu+1)},$$

получаем частное решение уравнения Бесселя

$$J_\nu(x) = \sum_{k=0}^{+\infty} (-1)^k \frac{\left(\frac{x}{2}\right)^{\nu+2k}}{k! \, \Gamma(\nu+k+1)}, \qquad (50)$$

где $J_\nu(x)$ — функция Бесселя (или цилиндрическая функция) индекса ν первого рода.

Если $\nu = n$, где $n \geqslant 0$ — целое, то

$$J_n(x) = \sum_{k=0}^{+\infty} (-1)^k \frac{\left(\frac{x}{2}\right)^{n+2k}}{k! \, (n+k)!}. \qquad (51)$$

Если n — четное, то функция Бесселя $J_n(x)$ — четная функция, если n — нечетное, то $J_n(x)$ — нечетная функция.

В частности, функция Бесселя нулевого индекса первого рода

$$J_0(x) = \sum_{k=0}^{+\infty} (-1)^k \frac{\left(\frac{x}{2}\right)^{2k}}{(k!)^2} = 1 - \frac{x^2}{4} + \frac{x^4}{64} - \ldots$$

Ее график имеет вид, представленный на рис. 37.

Формулы (50) и (51) определяют функции Бесселя в случае, когда индекс ν не равен целому отрицательному числу. Если же $\nu = -n$, где $n > 0$ — целое, то функция Бесселя $J_{-n}(x)$ определяется по такой формуле:

$$J_{-n}(x) = (-1)^n \sum_{k=0}^{+\infty} (-1)^k \frac{\left(\frac{x}{2}\right)^{n+2k}}{k! \, (n+k)!}. \qquad (52)$$

Очевидно, $J_{-n}(x) = (-1)^n \cdot J_n(x)$. Если индекс ν не является целым, то функции Бесселя $J_\nu(x)$ и $J_{-\nu}(x)$ оказываются линейно-независимыми, и из них можно построить общее решение уравнения Бесселя в виде

$$y = C_1 J_\nu(x) + C_2 J_{-\nu}(x).$$

В случае, когда n — целое, функции $J_n(x)$ и $J_{-n}(x)$ — линейно-зависимы, и из них нельзя построить общее решение уравнения Бесселя. Чтобы построить общее решение уравнения Бесселя в случае, когда $\nu = n$ — целое, надо найти еще одно решение этого уравнения, линейно-независимое от найденного выше частного решения $J_n(x)$. Таким решением является бесселева функция второго

рода или функция Неймана (Вебера) $Y_n(x)$. Иногда эта функция обозначается $N_n(x)$.

Функция Неймана $Y_\nu(x)$ для целого индекса ν определяется через $J_\nu(x)$ и $J_{-\nu}(x)$ по формуле

$$Y_\nu(x) = \frac{J_\nu(x) \cdot \cos \nu\pi - J_{-\nu}(x)}{\sin \nu\pi}.$$

Тогда $Y_n(x) = \lim\limits_{\nu \to n} Y_\nu(x)$, где n — целое. Функцию Неймана $Y_n(x)$ можно записать в виде некоторого степенного ряда.

На рис. 38 показан вид графика функции $Y_0(x)$.

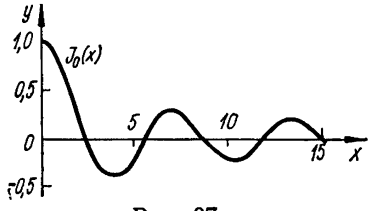

Рис. 37

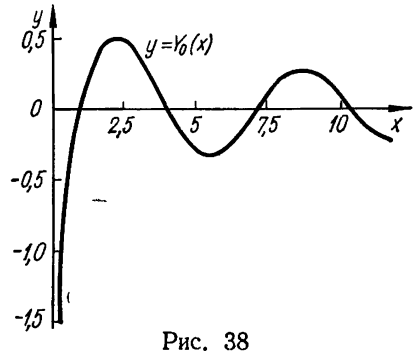

Рис. 38

Таблицы значений функций Бесселя и функций Неймана нулевого и первого порядка имеются в справочниках по математике.

Общее решение уравнения Бесселя целого индекса имеет вид

$$y = C_1 J_n(x) + C_2 Y_n(x).$$

Остановимся кратко еще на трех важных вопросах о функциях Бесселя.

1. Формулы приведения для функций Бесселя. Из формулы (50) находим

$$\frac{J_\nu(x)}{x^\nu} = \frac{1}{2^\nu} \sum_{k=0}^{+\infty} \frac{(-1)^k \left(\frac{x}{2}\right)^{2k}}{k!\,\Gamma(\nu + k + 1)}.$$

Тогда

$$\frac{d}{dx}\left(\frac{J_\nu(x)}{x^\nu}\right) = \frac{1}{2^\nu} \sum_{k=1}^{+\infty} \frac{(-1)^k \left(\frac{x}{2}\right)^{2k-1}}{(k-1)!\,\Gamma(\nu + k + 1)}.$$

Заменим здесь индекс суммирования k на $l + 1$. Получим

$$\frac{d}{dx}\left(\frac{J_\nu(x)}{x^\nu}\right) = \frac{1}{2^\nu} \sum_{l=0}^{+\infty} \frac{(-1)^{l+1} \left(\frac{x}{2}\right)^{2l+1}}{l!\,\Gamma(\nu + l + 2)} =$$

$$= -\frac{x}{2^{\nu+1}} \sum_{l=0}^{+\infty} \frac{(-1)^l \left(\frac{x}{2}\right)^{2l}}{l!\,\Gamma(\nu + l + 2)} = -x \frac{J_{\nu+1}(x)}{x^{\nu+1}}.$$

Таким образом,

$$\frac{d}{xdx}\left(\frac{J_\nu(x)}{x^\nu}\right) = -\frac{J_{\nu+1}(x)}{x^{\nu+1}}. \tag{53}$$

Отсюда, в частности, следует, что

$$J_0'(x) = -J_1(x).$$

Эта формула называется формулой повышения индекса бесселевой функции. Существует и формула понижения индекса:

$$\frac{d}{xdx}(x^\nu J_\nu(x)) = x^{\nu-1}J_{\nu-1}(x). \tag{54}$$

Применяя формулы (53) и (54), получаем такое равенство:

$$J_{\nu+1}(x) = \frac{2\nu}{x}J_\nu(x) - J_{\nu-1}(x). \tag{55}$$

Полагая здесь $\nu \in N$, можно последовательно выразить функции Бесселя целых индексов через функции Бесселя $J_0(x)$ и $J_1(x)$. Например,

$$J_2(x) = \frac{2}{x}J_1(x) - J_0(x).$$

2. Асимптотическая формула для функций Бесселя при больших значениях $|x|$. Если в уравнении Бесселя

$$y'' + \frac{1}{2}y' + \left(1 - \frac{\nu^2}{x^2}\right)y = 0$$

положить $y = \frac{z}{\sqrt{x}}$, где $z(x)$ — новая неизвестная функция, то оно примет вид

$$z'' + z - \left(\frac{\nu^2 - \frac{1}{4}}{x^2}\right)z = 0.$$

При больших $|x|$ это уравнение можно приближенно заменить уравнением

$$z'' + z = 0.$$

Тогда $z = C_1 \cos x + C_2 \sin x = A\cos(x + \alpha)$.

Следовательно,

$$y = \frac{A}{\sqrt{x}}\cos(x + \alpha).$$

При соответствующем выборе постоянных $A = A_\nu$ и $\alpha = \alpha_\nu$ получим приближенное выражение функции Бесселя

$$J_\nu(x) = \frac{A_\nu}{\sqrt{x}}\cos(x + \alpha_\nu).$$

В частности, $J_0(x) \approx \sqrt{\frac{2}{\pi x}}\cos\left(x - \frac{\pi}{4}\right)$.

В теории бесселевых функций доказывается, что эти функции имеют бесконечное множество нулей, причем расстояние между двумя последовательными нулями стремится к π при неограниченном возрастании $|x|$.

Имеются таблицы нулей функций Бесселя различных индексов. Например, положительные нули функции $J_0(x)$ будут в точках: 2,405; 5,520; 8,654; 11,792 и т. д. (см. рис. 37).

3. Ортогональность функций Бесселя. Функции $f_1(x)$ и $f_2(x)$ по определению ортогональны в $[a; b]$ с весом $\varphi(x)$, если

$$\int_a^b \varphi(x) \cdot f_1(x) \cdot f_2(x)\, dx = 0.$$

Пусть α и β — два различных по модулю нуля функции Бесселя $J_\nu(x)$ неотрицательного индекса ν.

Тогда доказывается, что

$$\int_0^1 x \cdot J_\nu(\alpha x) \cdot J_\nu(\beta x)\, dx = 0,$$

т. е. функции $J_\nu(\alpha x)$ и $J_\nu(\beta x)$ ортогональны с весом x в $[0; 1]$.

Свойство ортогональности функций Бесселя используется при разложении функций в ряд Фурье — Бесселя. Заметим, что существуют интегральные представления функций Бесселя целого индекса:

$$J_n(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(x \sin t - nt)}\, dt,$$

или

$$J_n(x) = \frac{1}{\pi} \int_0^{\pi} \cos(x \sin t - nt)\, dt.$$

В частности, при $n = 0$

$$J_0(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ix \sin t}\, dt,$$

или

$$J_0(x) = \frac{1}{\pi} \int_0^{\pi} \cos(x \sin t)\, dt.$$

Легко проверить, что

$$e^{ix \sin t} = \sum_{n=-\infty}^{+\infty} J_n(x) \cdot e^{int}.$$

Следовательно, функции Бесселя $J_n(x)$ являются коэффициентами ряда Фурье функции $e^{ix \sin t}$ в $[-\pi; \pi]$.

Функции Бесселя и Неймана от комплексного аргумента используются при решении многих задач. Например, в электротехнике при изучении поверхностного эффекта в цилиндрическом проводе кругового сечения встречаются функции $J_0(x)$ и $J_1(x)$, где $x = |x|\sqrt{-1}$. Таблицы функций $J_0(x)$ и $J_1(x)$ имеются обычно в любом учебнике по теоретической электротехнике.

Дельта-функция. Пусть $\varepsilon > 0$. Рассмотрим непрерывную функцию $\delta_\varepsilon(x)$, удовлетворяющую свойствам:

1) $\delta_\varepsilon(x) = 0$ при $x \leqslant -\varepsilon$ и при $x \geqslant \varepsilon$;
2) $\delta_\varepsilon(x) \geqslant 0$ при $-\varepsilon < x < \varepsilon$;
3) $\int\limits_{-\infty}^{\infty} \delta_\varepsilon(x)\,dx = \int\limits_{-\varepsilon}^{\varepsilon} \delta_\varepsilon(x)\,dx = 1.$

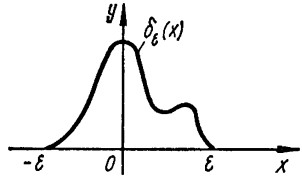

Рис. 39

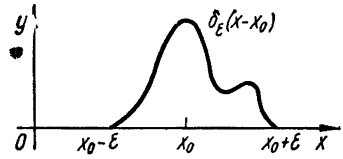

Рис. 40

Очевидно, при заданном ε можно построить сколь угодно много функций $\delta_\varepsilon(x)$, удовлетворяющих указанным выше условиям, причем среди этих функций могут быть и функции, имеющие производные любого порядка. График одной из функций $\delta_\varepsilon(x)$ приведен на рис. 39.

Рассмотрим также функцию $\delta_\varepsilon(x - x_0)$, график которой получится смещением графика функции $\delta_\varepsilon(x)$ на x_0 в направлении оси Ox (рис. 40).

Очевидно, функция $\delta_\varepsilon(x - x_0)$ удовлетворяет условиям:

1) $\delta_\varepsilon(x - x_0) = 0$ при $x < x_0 - \varepsilon$ и при $x > x_0 + \varepsilon$;
2) $\delta_\varepsilon(x - x_0) \geqslant 0$ при $x_0 - \varepsilon < x < x_0 + \varepsilon$;
3) $\int\limits_{-\infty}^{+\infty} \delta_\varepsilon(x - x_0)\,dx = \int\limits_{x_0-\varepsilon}^{x_0+\varepsilon} \delta_\varepsilon(x - x_0)\,dx = 1.$

Чем меньше ε, тем выше должен подниматься график функций $\delta_\varepsilon(x)$ и $\delta_\varepsilon(x - x_0)$, так как площадь под кривой на участке длиной 2ε согласно свойству 3) должна быть всегда равна 1. Точка x_0 называется центром импульса.

Пусть $\varepsilon \to 0$. Тогда говорят, что функция $\delta_\varepsilon(x - x_0)$ стремится к дельта-функции (импульсной функции) $\delta(x - x_0)$. Дельта-функция обладает следующими свойствами:

1) $\delta(x - x_0) = 0$ при $x \neq x_0$;
2) $\delta(x_0) = \infty$;
3) $\int\limits_{-\infty}^{+\infty} \delta(x - x_0)\,dx = 1.$

Ясно, что это не совсем обычная функция; ее следует рассматривать лишь в соответствии с указанным выше определением. Оказывается, что для всякой непрерывной функции $f(x)$

$$\int_a^b \delta(x - x_0) \cdot f(x)\, dx = f(x_0) \quad (a < x_0 < b).$$

При этом пределы интегрирования a и b могут быть бесконечными.

Дельта-функция $\delta(x - x_0)$ часто встречается во многих приложениях, в частности, когда наряду с непрерывно распределенными величинами (масса, заряд, тепловые источники, механический импульс и т. п.) приходится иметь дело с сосредоточенными величинами (точечная масса, точечный заряд и т. п.).

Существуют различные представления дельта-функции. В частности, ее можно представить в виде ряда Фурье:

$$\delta(x - x_0) = \sum_{n=1}^{\infty} \frac{2}{l} \sin \frac{n\pi x}{l} \cdot \sin \frac{n\pi x_0}{l}$$

или в виде интеграла Фурье:

$$\delta(x - x_0) = \frac{1}{\pi} \int_0^{\infty} \cos \lambda (x - x_0)\, d\lambda.$$

§ 16. РЕШЕНИЕ ТРЕХМЕРНОГО ОДНОРОДНОГО ВОЛНОВОГО УРАВНЕНИЯ

Существуют различные методы нахождения частных решений трехмерного однородного волнового уравнения

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} - \frac{1}{a^2} \cdot \frac{\partial^2 u}{\partial t^2} = 0.$$

Это уравнение кратко записывают так:

$$\Delta u - \frac{1}{a^2} \frac{\partial^2 u}{\partial t^2} = 0, \tag{56}$$

где $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ — оператор Лапласа.

Найдем решение уравнения (56) типа сферической волны. С этой целью рассмотрим семейство концентрических сфер

$$(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = \rho^2$$

с центром в фиксированной точке (x_0, y_0, z_0). Параметром этого семейства является радиус сфер

$$\rho = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}.$$

Решение уравнения (56) будем искать в виде

$$u = u(\rho, t).$$

Тогда фазовыми поверхностями решения будут указанные выше концентрические сферы.

Запишем выражение для оператора Лапласа в сферических координатах (см. гл. V):

$$\Delta u = \frac{1}{\rho^2 \sin\theta}\left(\frac{\partial}{\partial\rho}\left(\rho^2 \cdot \sin\theta \cdot \frac{\partial u}{\partial\rho}\right) + \frac{\partial}{\partial\theta}\left(\sin\theta \cdot \frac{\partial u}{\partial\theta}\right) + \right.$$

$$\left. + \frac{\partial}{\partial\varphi}\left(\frac{1}{\sin\theta} \cdot \frac{\partial u}{\partial\varphi}\right)\right).$$

В нашем случае u не зависит ни от θ, ни от φ. Поэтому $\frac{\partial u}{\partial\theta} = 0$, $\frac{\partial u}{\partial\varphi} = 0$ и

$$\Delta u = \frac{1}{\rho^2 \sin\theta} \cdot \frac{\partial}{\partial\rho}\left(\rho^2 \sin\theta \cdot \frac{\partial u}{\partial\rho}\right) = \frac{1}{\rho^2} \cdot \frac{\partial}{\partial\rho}\left(\rho^2 \cdot \frac{\partial u}{\partial\rho}\right) =$$

$$= \frac{1}{\rho^2}\left(2\rho \cdot \frac{\partial u}{\partial\rho} + \rho^2 \cdot \frac{\partial^2 u}{\partial\rho^2}\right) = \frac{\partial^2 u}{\partial\rho^2} + \frac{2}{\rho} \cdot \frac{\partial u}{\partial\rho}.$$

Легко проверить, что

$$\frac{\partial^2 u}{\partial\rho^2} + \frac{2}{\rho} \cdot \frac{\partial u}{\partial\rho} = \frac{1}{\rho} \cdot \frac{\partial^2 (\rho u)}{\partial\rho^2}.$$

Следовательно,

$$\Delta u = \frac{1}{\rho} \cdot \frac{\partial^2 (\rho u)}{\partial\rho^2}.$$

Теперь уравнение (56) примет вид

$$\frac{1}{\rho} \cdot \frac{\partial^2 (\rho u)}{\partial\rho^2} - \frac{1}{a^2} \cdot \frac{\partial^2 u}{\partial t^2} = 0,$$

или

$$\frac{\partial^2 (\rho u)}{\partial\rho^2} - \frac{1}{a^2} \cdot \frac{\partial^2 (\rho u)}{\partial t^2} = 0.$$

Таким образом, получили однородное волновое уравнение относительно функции ρu.

Общее решение этого уравнения может рассматриваться как суперпозиция двух встречных бегущих волн, т. е.

$$\rho u = \Phi(\rho - at) + \Psi(\rho + at).$$

Тогда решение уравнения (56) типа сферической бегущей волны запишется так:

$$u = \frac{1}{\rho}\Phi(\rho - at) + \frac{1}{\rho}\Psi(\rho + at).$$

Здесь Φ и Ψ — произвольные функции.

Наиболее часто встречается сферическая бегущая волна

$$\frac{1}{\rho}\Phi(\rho - at).$$

Она является затухающей, так как при неограниченном увеличении времени t на расширяющейся сферической фазовой поверхности $\rho = a(t + t_0)$ значение волновой функции $\frac{\Phi(at_0)}{a(t+t_0)}$ стремится к нулю.

Построим теперь решение уравнения (56) в виде суперпозиции прямых расходящихся затухающих сферических бегущих волн. Предположим, что в области V задана непрерывная функция $g(x, y, z)$.

Разобьем область V произвольным способом на n частичных областей Δv_k, в каждой из них выберем по точке (ξ_k, η_k, ζ_k).

Функция $$u_k = \frac{\Phi(\rho_k - at)}{\rho_k},$$

где $\rho_k = \sqrt{(x-\xi_k)^2 + (y-\eta_k)^2 + (z-\zeta_k)^2}$, является решением уравнения (56).

Так как уравнение (56) линейное и однородное, то

$$\sum_{k=1}^{n} g(\xi_k, \eta_k, \zeta_k) \cdot \frac{\Phi(\rho_k - at)}{\rho_k} \cdot \Delta v_k$$

также является его решением (как линейная комбинация решений $\frac{\Phi(\rho_k - at)}{\rho_k}$ с постоянными коэффициентами $g(\xi_k, \eta_k, \zeta_k) \cdot \Delta v_k$). Можно показать, что функция

$$u = \lim_{\substack{n\to\infty \\ (\max \Delta v_k \to 0)}} \sum_{k=1}^{n} g(\xi_k, \eta_k, \zeta_k) \cdot \frac{\Phi(\rho_k - at)}{\rho_k} \cdot \Delta v_k =$$

$$= \iiint_V g(\xi, \eta, \zeta) \cdot \frac{\Phi(\rho - at)}{\rho} \cdot d\xi \cdot d\eta \cdot d\zeta, \qquad (57)$$

где $\rho = \sqrt{(x-\xi)^2 + (y-\eta)^2 + (z-\zeta)^2}$, также является решением уравнения (56).

Перейдем в тройном интеграле (57) к сферическим координатам ρ, φ, θ, положив $\xi - x = \rho \sin\theta \cdot \cos\varphi$, $\eta - y = \rho \sin\theta \cdot \sin\varphi$, $\zeta - z = \rho\cos\theta$.

Тогда $\xi = x + \rho\sin\theta\cdot\cos\varphi$, $\eta = y + \rho\sin\theta\cdot\sin\varphi$, $\zeta = z + \rho\cos\theta$,

$$u = \int_0^{2\pi} d\varphi \int_0^{\pi} \sin\theta d\theta \int_0^{\rho(\varphi,\theta)} g(x + \rho\sin\theta\cdot\cos\varphi,\ y + \rho\sin\theta\cdot\sin\varphi,\ z + \rho\cos\theta) \cdot \frac{\Phi(\rho - at)}{\rho} \cdot \rho^2 d\rho =$$

$$= \int_0^{2\pi} d\varphi \int_0^{\pi} \sin\theta d\theta \int_0^{\rho(\varphi,\theta)} g(x + \rho\sin\theta\cdot\cos\varphi,\ y + \rho\sin\theta\cdot\sin\varphi,\ z + \rho\cos\theta) \cdot \rho \cdot \Phi(\rho - at)\, d\rho. \qquad (58)$$

Теперь вместо произвольной функции $\Phi(\rho - at)$ возьмем функцию $\frac{1}{4\pi a}\delta(\rho - at)$, где $\delta(\rho - at)$ — дельта-функция с центром импульса $\rho_0 = at$ (t — фиксировано).

По основному свойству дельта-функции

$$\int_0^{\rho(\varphi,\theta)} g(x + \rho\sin\theta\cos\varphi,\ y + \rho\sin\theta\sin\varphi,\ z + \rho\cos\theta)\,\rho\delta(\rho - at)\,d\rho =$$

$$= g(x + at\sin\theta\cos\varphi,\ y + at\sin\theta\sin\varphi,\ z + at\cos\theta)\,at$$

$$\left(\int_0^{\rho} \tilde{g}(\rho)\,\delta(\rho - \rho_0)\,d\rho = \tilde{g}(\rho_0)\right);$$

в нашем случае $\tilde{g}(\rho) = g(\rho, \theta, \varphi)\cdot\rho$, $\rho_0 = at$. Учитывая это, формулу (58) записываем в виде:

$$u = \frac{t}{4\pi}\int_0^{\pi}\sin\theta d\theta\int_0^{2\pi} g(x + at\sin\theta\cos\varphi,\ y + at\sin\theta\sin\varphi,\ z +$$

$$+ at\cos\theta)\,d\varphi. \qquad (59)$$

Здесь мы получили решение уравнения (56), содержащее произвольную функцию $g(x, y, z)$. Из выражения (59) видно, что

$$u|_{t=0} = 0. \qquad (60)$$

Найдем

$$\frac{\partial u}{\partial t} = \frac{1}{4\pi}\int_0^{\pi}\sin\theta d\theta\int_0^{2\pi} g(x + at\sin\theta\cos\varphi,\ y + at\sin\theta\sin\varphi,\ z +$$

$$+ at\cos\theta)\,d\theta + \frac{t}{4\pi}\frac{d}{dt}\left(\int_0^{\pi}\sin\theta d\theta\int_0^{2\pi} g(x + at\sin\theta\cos\varphi,\ y +\right.$$

$$\left. + at\sin\theta\sin\varphi,\ z + at\cos\theta)\,d\theta\right).$$

Полагая здесь $t = 0$, получаем

$$\left.\frac{\partial u}{\partial t}\right|_{t=0} = \frac{1}{4\pi}\int_0^{\pi}\sin\theta d\theta\int_0^{2\pi} g(x, y, z)\,d\theta = \frac{2\pi}{4\pi}g(x, y, z)\int_0^{\pi}\sin\theta d\theta =$$

$$= \frac{1}{2}g(x, y, z)(-\cos\theta)\Big|_0^{\pi} = g(x, y, z).$$

Таким образом,

$$\left.\frac{\partial u}{\partial t}\right|_{t=0} = g(x, y, z). \qquad (61)$$

Функция u (59) применяется для решения задачи Коши, которая формулируется так:

Найти решение однородного волнового уравнения

$$\Delta u - \frac{1}{a^2} \cdot \frac{\partial^2 u}{\partial t^2} = 0,$$

удовлетворяющее начальным условиям

$$u|_{t=0} = g_1(x, y, z), \quad \left.\frac{\partial u}{\partial t}\right|_{t=0} = g_2(x, y, z), \tag{62}$$

где $g_1(x, y, z)$ *и* $g_2(x, y, z)$ *— заданные (непрерывные) функции.*

Подставляя в формулу (59) вместо произвольной функции $g(x, y, z)$ заданные функции $g_1(x, y, z)$ и $g_2(x, y, z)$, находим два частных решения уравнения (56):

$$u_1 = \frac{t}{4\pi} \int_0^{\pi} \sin\theta d\theta \int_0^{2\pi} g_1(x + at\sin\theta\cos\varphi,\ y + at\sin\theta\sin\varphi,\ z + at\cos\theta)\, d\theta,$$

$$u_2 = \frac{t}{4\pi} \int_0^{\pi} \sin\theta d\theta \int_0^{2\pi} g_2(x + at\sin\theta\cos\varphi,\ y + at\sin\theta\sin\varphi,\ z + at\cos\theta)\, d\theta.$$

Непосредственно подстановкой в уравнение (56) можно проверить, что функция $\frac{\partial u_1}{\partial t}$ является также решением этого уравнения. Последнее справедливо для любого решения этого уравнения, в частности и для u_2.

Рассмотрим функцию

$$u = u_2 + \frac{\partial u_1}{\partial t} =$$

$$= \frac{t}{4\pi} \int_0^{\pi} \sin\theta d\theta \int_0^{2\pi} g_2 d\theta + \frac{\partial}{\partial t}\left\{\frac{1}{4\pi} \int_0^{\pi} \sin\theta d\theta \int_0^{2\pi} g_1 d\theta\right\} =$$

$$= \frac{t}{4\pi} \int_0^{\pi} \sin\theta d\theta \int_0^{2\pi} g_2 d\theta + \frac{t}{4\pi} \int_0^{\pi} \sin\theta d\theta \int_0^{2\pi} g_1 d\theta +$$

$$+ \frac{t}{4\pi} \cdot \frac{\partial}{\partial t} \int_0^{\pi} \sin\theta d\theta \int_0^{2\pi} g_1 d\theta. \tag{63}$$

Так как u_2 и $\frac{\partial u_1}{\partial t}$ являются решениями уравнения (56), то в силу его линейности и однородности функция $u = u_2 + \frac{\partial u_1}{\partial t}$ также будет его решением. Проверка показывает, что эта функция удовлетворяет и начальным условиям (62), т. е. представляет собой решение задачи Коши для уравнения (56). Проверим выполнение

первого из начальных условий (62). По формуле (63)

$$u|_{t=0} = \frac{1}{4\pi}\int_0^{\pi} \sin\theta d\theta \int_0^{2\pi} g_1(x, y, z)\, d\theta = g_1(x, y, z)$$

(аналогичный интеграл был вычислен в формуле (61)).

Таким образом, функция u (63) удовлетворяет первому из начальных условий (62). Найдем

$$\frac{\partial u}{\partial t} = \frac{\partial u_2}{\partial t} + \frac{\partial^2 u_1}{\partial t^2},$$

тогда

$$\frac{\partial u}{\partial t}\Big|_{t=0} = \frac{\partial u_2}{\partial t}\Big|_{t=0} + \frac{\partial^2 u_1}{\partial t^2}\Big|_{t=0}.$$

По формуле (61) имеем

$$\frac{\partial u_2}{\partial t}\Big|_{t=0} = g_2(x, y, z).$$

Для нахождения $\frac{\partial^2 u_1}{\partial t^2}$ учтем, что u_1 удовлетворяет уравнению (56), следовательно,

$$\frac{\partial^2 u_1}{\partial x^2} + \frac{\partial^2 u_1}{\partial y^2} + \frac{\partial^2 u_1}{\partial z^2} - \frac{1}{a^2}\cdot\frac{\partial^2 u_1}{\partial t^2} = 0.$$

Отсюда получим, что

$$\frac{\partial^2 u_1}{\partial t^2} = a^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)u_1 =$$

$$= \frac{a^2 t}{4\pi}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\int_0^{\pi}\sin\theta d\theta \int_0^{2\pi} g_1 d\theta.$$

Полагая в этом равенстве $t = 0$, получаем

$$\frac{\partial^2 u_1}{\partial t^2}\Big|_{t=0} = 0.$$

Тогда $\frac{\partial u}{\partial t}\Big|_{t=0} = \frac{\partial u_2}{\partial t}\Big|_{t=0} + \frac{\partial^2 u_1}{\partial t^2}\Big|_{t=0} = g_2(x, y, z)$.

Таким образом, для функции u (63) выполняется и второе начальное условие (62).

Формула

$$u = \frac{t}{4\pi}\int_0^{\pi}\sin\theta d\theta \int_0^{2\pi} g_2(x + at\sin\theta\cos\varphi,\ y + at\sin\theta\sin\varphi,\ z +$$

$$+ at\cos\theta)\, d\theta + \frac{\partial}{\partial t}\left\{\frac{1}{4\pi}\int_0^{\pi}\sin\theta d\theta \int_0^{2\pi} g_1(x + at\sin\theta\cos\varphi,\ y +\right.$$

$$\left. + at\sin\theta\sin\varphi,\ z + at\cos\theta)\, d\theta\right\}$$

называется формулой Пуассона. Она дает решение задачи Коши для однородного трехмерного волнового уравнения (56). Формулу Пуассона иногда записывают в виде

$$u(x, y, z) = \frac{\partial}{\partial t}\left(\frac{t}{\sigma}\iint\limits_{\sigma} g_1(\xi, \eta, \zeta)\, d\sigma\right) + \frac{t}{\sigma}\iint\limits_{\sigma} g_2(\xi, \eta, \zeta)\, d\sigma, \quad (64)$$

где интегрирование выполняется по поверхности сферы σ с центром в точке (x, y, z) переменного радиуса $\rho = at$.

ГЛАВА IV

ОПЕРАЦИОННОЕ ИСЧИСЛЕНИЕ

Операционное (символическое) исчисление в настоящее время широко применяется в самых различных областях науки и техники. Особенно большую роль оно играет при исследовании переходных процессов в линейных физических системах электротехники, автоматики, радиотехники, механики и других отраслей знаний.

Современный математический аппарат операционного исчисления позволяет решать задачи, описываемые системами линейных дифференциальных уравнений (обыкновенных и с частными производными), разностными и дифференциально-разностными уравнениями, линейными дифференциальными уравнениями с переменными коэффициентами и некоторыми типами интегральных уравнений. Такая большая универсальность метода объясняется его эффективностью — возможностью получить решение наиболее простыми и экономными путями и средствами.

Строгое математическое обоснование и развитие операционное исчисление получило на базе общей теории функциональных (интегральных) преобразований и, в частности, преобразований Фурье и Лапласа.

§ 17. ПРЕОБРАЗОВАНИЕ ЛАПЛАСА

Оригинал и изображение. *Оригиналом называется всякая функция $f(t)$ действительного переменного t, удовлетворяющая следующим условиям:*

1) *$f(t)$ — однозначная, непрерывная или кусочно-непрерывная вместе со своими производными n-го порядка при $t \geqslant 0$;*

2) *$|f(t)|$ растет не быстрее, чем некоторая показательная функция, что означает существование таких постоянных положительных чисел M и S_0, не зависящих от t, при которых $|f(t)| \leqslant Me^{S_0 t}$ для всех $t \geqslant 0$;*

3) $f(t) \equiv 0$ при $t < 0$.

Класс функций-оригиналов будем называть к л а с с о м D, так что $f(t) \in D$.

И з о б р а ж е н и е м функции-оригинала $f(t)$ называется функция $F(p)$ комплексного переменного $p = \sigma + i\omega$, определяемая интегралом Лапласа

$$F(p) = \int_0^\infty e^{-pt} f(t)\, dt. \tag{1}$$

Интеграл Лапласа (1), называемый часто п р е о б р а з о в а н и е м Л а п л а с а функции $f(t)$, несобственный. Он равномерно сходится, если функция $f(t)$ удовлетворяет наложенным на оригинал условиям. Действительно, при этом имеем

$$\int_0^\infty |f(t)|\, e^{-pt}\, dt \leqslant M \int_0^\infty e^{-pt} e^{S_0 t}\, dt = M \int_0^\infty e^{(S_0 - p)t}\, dt =$$

$$= \frac{M}{S_0 - p} e^{(S_0 - p)} \Big|_0^\infty = \frac{M}{p - S_0}$$

(при $p > S_0$, что обычно понимается как $\operatorname{Re} p > S_0$).

Функцию $F(p)$ называют и з о б р а ж е н и е м ф у н к ц и и $f(t)$ п о Л а п л а с у.

Соответствие между изображением и оригиналом будем обозначать знаком $\doteqdot$ и записывать

$$F(p) \doteqdot f(t) \text{ или } f(t) \doteqdot F(p). \tag{2}$$

Замечания

1. Из определения интеграла Лапласа следует, что условие 3), налагаемое на оригинал, чисто формальное. Если говорят об изображении функции $f(t)$, не равной тождественно нулю при $t < 0$, то имеют в виду лишь изображение функции $f(t) \cdot \sigma(t)$, где $\sigma(t)$ — е д и н и ч н а я ф у н к ц и я

$$\sigma(t) = \begin{cases} 0 \text{ при } t < 0, \\ 1 \text{ при } t \geqslant 0. \end{cases}$$

Действительно, записанное произведение совпадает с функцией $f(t)$ при $t \geqslant 0$ и равно тождественно нулю при $t < 0$, т. е. это функция, принадлежащая классу D, и для нее можно найти изображение. Например, говоря об изображении функции $\sin \omega t$, предполагают изображение функции

$$\varphi(t) = \sin \omega t \cdot \sigma(t) = \begin{cases} 0 & \text{при } t < 0, \\ \sin \omega t & \text{при } t \geqslant 0. \end{cases}$$

2. Помимо изображения по Лапласу (1) применяется также изображение функции по К а р с о н у (или по Х е в и с а й д у) $\Phi(p) = p \int_0^\infty e^{-pt} f(t)\, dt \doteqdot f(t)$, отличающееся от преобразования Лапласа множителем p. В последнее время в технической литературе все чаще пользуются изображением по Лапласу. Это объясняется наличием наглядной связи между операторным (символическим) методом и гармоническим анализом, вносящей физический смысл в понятие изображения (изображение по Лапласу $F(\sigma + i\omega)$ — это спектральная функция по отношению к затухающей функции $e^{-\sigma t} f(t)$, для которой переменная ω является частотой) [1].

[1] См.: В. С. М а р т ы н е н к о. Операционное исчисление. К., «Вища школа», 1973.

3. Для обозначения соответствия между оригиналом и изображением употребляются, кроме принятого выше, еще и другие символы: $\div$, $\Leftrightarrow$, $\bullet - \circ$, $\rightarrow$, $F(p) = L\{f(t)\}$.

Свойства изображений. Изображения обладают рядом свойств, которые используются при расчетах с помощью операционного исчисления.

Рассмотрим эти свойства, причем первые два из них примем без доказательства.

1. ***Теорема единственности.*** *Если два изображения* $F(p)$ и $\Phi(p)$ *совпадают, то совпадают между собой и соответствующие оригиналы во всех точках, за исключением, быть может, точек разрыва, т. е. если* $F(p) \div f(t)$ *и* $\Phi(p) \div \varphi(t)$ *и при этом* $F(p) = \Phi(p)$, *то* $f(t) \equiv \varphi(t)$ *во всех точках непрерывности.*

2. ***Теорема об аналитичности изображения.*** *Всякое изображение* $F(p)$ *при* $\operatorname{Re} p > S_0$ *является аналитической функцией, т. е. может быть разложено в степенной ряд и, следовательно, неограниченное число раз интегрируемо и дифференцируемо в области сходимости ряда.*

3. ***Свойства линейности.*** *Если* $f(t) = C_1 f_1(t) + C_2 f_2(t)$ *и при этом* $f_1(t) \div F_1(p)$, $f_2(t) \div F_2(p)$, $f(t) \div F(p)$, *то*

$$f(t) \div F(p) = C_1 F_1(p) + C_2 F_2(p).$$

Доказательство. По определению изображения имеем

$$F(p) = \int_0^\infty f(t)\, e^{-pt}\, dt = \int_0^\infty (C_1 f_1(t) + C_2 f_2(t))\, e^{-pt}\, dt =$$

$$= C_1 \int_0^\infty f_1(t)\, e^{-pt} + C_2 \int_0^\infty f_2(t)\, e^{-pt}\, dt = C_1 F_1(p) + C_2 F_2(p).$$

4. *Всякое изображение* $F(p)$ *функции* $f(t)$ *при* $p \to \infty$ *стремится к нулю.*

Доказательство. Если $f(t)$ — оригинал, то можно выбрать такие постоянные $M > 0$ и $S_0 > 0$, что $|f(t)| < Me^{S_0 t}$ $(t > 0)$.

Тогда при $p > S_0$ получим

$$|F(p)| = \left| \int_0^\infty f(t)\, e^{-pt}\, dt \right| \leqslant \int_0^\infty |f(t)|\, e^{-pt}\, dt \leqslant M \int_0^\infty e^{S_0 t} e^{-pt}\, dt =$$

$$= M \int_0^\infty e^{-(p-S_0)t}\, dt = \frac{M}{p - S_0} \to 0 \text{ при } p \to \infty.$$

Следовательно, $\lim\limits_{p \to \infty} |F(p)| = 0$.

Отсюда, в частности, следует, что константы и многочлены от p с положительными степенями не могут быть изображениями.

Изображения некоторых простейших функций. Найдем (непосредственным интегрированием и используя свойство линей-

ности) изображения функций $\sigma(t)$, $e^{-\alpha t}$, $e^{\alpha t}$, $\operatorname{sh}\alpha t$, $\operatorname{ch}\alpha t$, $\sin\omega t$, $\cos\omega t$, $1-e^{-\alpha t}$.

Эти функции принадлежат классу оригиналов, так как каждая из них удовлетворяет условиям 1) — 3).

Изображение единичной функции

$$\sigma(t)=\begin{cases}0 \text{ при } t<0,\\ 1 \text{ при } t\geqslant 0\end{cases}$$

(рис. 41).

Найдем $\int\limits_0^\infty \sigma(t)e^{-pt}dt=\int\limits_0^\infty e^{-pt}dt=\frac{1}{p}e^{-pt}\Big|_0^\infty=\frac{1}{p}$ $(p>0)$.

Итак, $\sigma(t)\doteqdot\frac{1}{p}$ (иногда пишут $1\doteqdot\frac{1}{p}$).

Изображения показательных функций.

$$e^{-\alpha t}\doteqdot\int\limits_0^\infty e^{-\alpha t}e^{-pt}dt=\int\limits_0^\infty e^{-(p+\alpha)t}dt=\frac{-e^{-(p+\alpha)t}}{p+\alpha}\Big|_0^\infty=\frac{1}{p+\alpha},$$

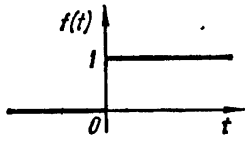

Рис. 41

$$e^{\alpha t}\doteqdot\int\limits_0^\infty e^{\alpha t}e^{-pt}dt=\int\limits_0^\infty e^{-(p-\alpha)t}dt=\frac{1}{p-\alpha}.$$

При $\alpha=i\omega$ получим

$$e^{-i\omega t}\doteqdot\frac{1}{p+i\omega} \text{ и } e^{i\omega t}\doteqdot\frac{1}{p-i\omega}.$$

Изображения гиперболических функций.

$$\operatorname{ch}\alpha t=\frac{1}{2}(e^{\alpha t}+e^{-\alpha t})\doteqdot\frac{1}{2}\left(\frac{p}{p-\alpha}+\frac{1}{p+\alpha}\right)=$$

$$=\frac{p+\alpha+p-\alpha}{2(p^2-\alpha^2)}=\frac{p}{p^2-\alpha^2},$$

$$\operatorname{sh}\alpha t=\frac{1}{2}(e^{\alpha t}-e^{-\alpha t})\doteqdot\frac{1}{2}\left(\frac{1}{p-\alpha}-\frac{1}{p+\alpha}\right)=$$

$$=\frac{p+\alpha-p+\alpha}{2(p^2-\alpha^2)}=\frac{\alpha}{p^2-\alpha^2}.$$

Изображения тригонометрических функций.

$$\cos\omega t=\frac{1}{2}(e^{i\omega t}+e^{-i\omega t})\doteqdot\frac{1}{2}\left(\frac{1}{p-i\omega}+\frac{1}{p+i\omega}\right)=$$

$$=\frac{p+i\omega+p-i\omega}{2(p^2+\omega^2)}=\frac{p}{p^2+\omega^2},$$

$$\sin\omega t=\frac{1}{2i}(e^{i\omega t}-e^{-i\omega t})\doteqdot\frac{1}{2i}\left(\frac{1}{p-i\omega}-\frac{1}{p+i\omega}\right)=$$

$$=\frac{p+i\omega-p+i\omega}{2i(p^2+\omega^2)}=\frac{\omega}{p^2+\omega^2}.$$

Изображение функции $1 - e^{-\alpha t}$.

$$1 - e^{-\alpha t} \doteqdot \frac{1}{p} - \frac{1}{p+\alpha} = \frac{p+\alpha-p}{p(p+\alpha)} = \frac{\alpha}{p(p+\alpha)}.$$

Изображения производных от функции. Предположим, что функция $f(t)$ дифференцируема при $t > 0$, причем производная $f'(t) \in D$. Тогда сама функция $f(t)$ также принадлежит классу оригиналов.

Действительно, так как существует производная $f'(t)$, то $f(t)$ непрерывна при $t \geqslant 0$; кроме того, поскольку $f'(t) \in D$, постольку можно найти такие числа $M > 0$ и $S_0 > 0$, что

$$|f'(t)| < Me^{S_0 t}.$$

Так как $f(t) = \int_0^t f'(t)\,dt + f(0)$, то

$$|f(t)| \leqslant \int_0^t |f'(t)|\,dt + |f(0)| \leqslant \int_0^t Me^{S_0 t}\,dt + |f(0)| =$$

$$= \frac{M}{S_0}(e^{S_0 t} - 1) + |f(0)|.$$

Можно найти достаточно большое число $M_1 > 0$, что

$$\frac{M}{S_0}(e^{S_0 t} - 1) + |f(0)| \leqslant M_1 e^{S_0 t}.$$

Таким образом, функция $f(t)$ удовлетворяет всем условиям, предъявляемым к оригиналам, а потому принадлежит классу оригиналов.

Пусть теперь

$$f(t) \doteqdot \int_0^\infty f(t)\,e^{-pt}\,dt = F(p).$$

Тогда изображение производной получим, выполняя интегрирование по частям:

$$f'(t) \doteqdot \int_0^\infty f'(t)\,e^{-pt}\,dt = \left| \begin{array}{l} u = e^{-pt}, \quad du = -pe^{-pt}\,dt \\ dv = f'(t)\,dt, \quad v = f(t) \end{array} \right| =$$

$$= f(t)\,e^{-pt}\Big|_0^\infty + p\int_0^\infty f(t)\,e^{-pt}\,dt = -f(0) + pF(p).$$

Итак,

$$f'(t) \doteqdot pF(p) - f(0). \tag{3}$$

Пример. $f(t) = e^{\alpha t} \doteqdot \frac{1}{p-\alpha} = F(p)$. Найти изображение производной $f'(t)$.

Решение. Имеем

$$f'(t) \doteqdot pF(p) - f(0) = p\frac{1}{p-\alpha} - e^{\alpha t}\Big|_{t=0} = \frac{p}{p-\alpha} - 1,$$

откуда

$$f'(t) \doteqdot \frac{\alpha}{p-\alpha}.$$

Если предположить существование для функции $f(t)$ производных высших порядков и принадлежность их классу оригиналов, то легко найти их изображения, зная изображение первой производной. Действительно, обозначим $f'(t) = \varphi(t)$. Тогда $f''(t) = \varphi'(t)$.

Если $\varphi(t) \doteqdot \Phi(p)$, $f(t) \doteqdot F(p)$, то $\Phi(p) = pF(p) - f(0)$,

$$\varphi'(t) \doteqdot p\Phi(p) - \varphi(0) = p(pF(p) - f(0)) - f'(0),$$

$$\varphi'(t) \doteqdot p^2F(p) - pf(0) - f'(0),$$

т. е.

$$f''(t) \doteqdot p^2F(p) - pf(0) - f'(0). \tag{4}$$

Подобно этому найдем изображение третьей производной

$$f'''(t) \doteqdot p(p^2F(p) - pf(0) - f'(0)) - f''(0) =$$

$$= p^3F(p) - p^2f(0) - pf'(0) - f''(0),$$

а затем — изображение производной любого порядка:

$$f^{(n)}(t) \doteqdot p^nF(p) - p^{n-1}f(0) - p^{n-2}f'(0) - \ldots$$

$$\ldots - pf^{(n-2)}(0) - f^{(n-1)}(0). \tag{5}$$

Особенно просто выглядят формулы, определяющие изображения производных при нулевых начальных условиях.

Если $f(0) = f'(0) = \ldots = f^{(n-1)}(0) = 0$ и $f(t) \doteqdot F(p)$, то

$$f'(t) \doteqdot pF(p),\ f''(t) \doteqdot p^2F(p),\ \ldots,\ f^{(n)}(t) \doteqdot p^nF(p).$$

Примеры

1. $f(t) = \cos t$, $f(0) = 1$, $f'(0) = 0$. Найти изображение второй производной $f''(t)$.

Решение. Так как

$$f(t) \doteqdot F(p) = \frac{p}{p^2+1} \quad \left(\cos t \doteqdot \frac{p}{p^2+1}\right)$$

и

$$f''(t) \doteqdot p^2F(p) - pf(0) - f'(0),$$

то

$$f''(t) \doteqdot p^2 \cdot \frac{p}{p^2+1} - p \cdot 1 - 0 = \frac{p^3 - p(p^2+1)}{p^2+1} = \frac{-p}{p^2+1}.$$

2. $f(t) = e^{\alpha t}$. Найти изображение производной $f^{(n)}(t)$.

Решение. Так как $f(t) = e^{\alpha t} \doteqdot \frac{1}{p-\alpha}$ и, кроме того,

$$f(0) = e^{\alpha t}|_{t=0} = 1,\quad f^{(k)}(0) = \alpha^k e^{\alpha t}|_{t=0} = \alpha^k$$

$$(k = 1, 2, 3, \ldots, n),$$

то

$$f^{(n)}(t) \doteqdot p^n F(p) - p^{n-1} f(0) - p^{n-2} f'(0) - \ldots - p f^{(n-2)}(0) - f^{(n-1)}(0) =$$

$$= p^n \frac{1}{p-\alpha} - p^{n-1} \cdot 1 - p^{n-2}\alpha - p^{n-3}\alpha^2 - \ldots - p\alpha^{n-2} - \alpha^{n-1}.$$

Изображение интеграла от функции. Пусть $f(t)$ непрерывна и $f(t) \in D$. Требуется найти изображение интеграла $\varphi(t) = \int_0^t f(t)\,dt$. Очевидно,

$$\varphi'(t) = \left(\int_0^t f(\tau)\,d\tau\right)'_t = f(t).$$

По условию функция $f(t) = \varphi'(t)$ принадлежит классу оригиналов. Поэтому, согласно доказанному ранее, классу оригиналов принадлежит и функция

$$\varphi(t) = \int_0^t f(t)\,dt.$$

Пусть теперь $f(t) \doteqdot F(p)$ и $\varphi(t) \doteqdot \Phi(p)$.

Так как $f(t) = \varphi'(t) \doteqdot p\Phi(p) - \varphi(0)$, или $F(p) = p\Phi(p) - \varphi(0)$, то

$$\Phi(p) = \frac{F(p) + \varphi(0)}{p}.$$

Однако $\varphi(0) = \int_0^0 f(t)\,dt = 0$. Поэтому $\Phi(p) = \frac{1}{p} F(p)$.

Итак, если $f(t) \doteqdot F(p)$, то

$$\int_0^t f(\tau)\,d\tau \doteqdot \frac{1}{p} F(p). \qquad (6)$$

Пример. Известно, что $\cos t \doteqdot \frac{p}{p^2+1} = F(p)$. Найти изображение функции $\sin t = \int_0^t \cos t\,dt$.

Р е ш е н и е. Так как

$$\int_0^t f(\tau)\,d\tau \doteqdot \frac{F(p)}{p},$$

то

$$\sin t \doteqdot \frac{1}{p} \cdot \frac{p}{p^2+1} = \frac{1}{p^2+1}.$$

§ 18. ОСНОВНЫЕ ТЕОРЕМЫ ОПЕРАЦИОННОГО ИСЧИСЛЕНИЯ

Отыскание изображений большинства функций непосредственно с помощью интеграла Лапласа весьма громоздко. Изложенные ниже теоремы существенно облегчают решение задачи нахождения изображений большого числа разнообразных функций. Они позволяют также часто решать обратную задачу — отыскания оригиналов по изображениям. В результате этого математический аппарат операционного исчисления становится на практике очень удобным и эффективным средством решения многих задач.

Теорема о дифференцировании изображения (об умножении оригинала на аргумент). *Если* $f(t) \in D$ *и* $f(t) \doteqdot F(p)$, *то* $tf(t) \doteqdot -\frac{d}{dp} F(p)$.

Доказательство. Прежде всего отметим, что при указанном в теореме условии будет $tf(t) \in D$. Далее

$$tf(t) \doteqdot \int_0^\infty tf(t)\, e^{-pt}\, dt = -\int_0^\infty f(t) \frac{d}{dp}(e^{-pt})\, dt =$$

$$= -\int_0^\infty \frac{d}{dp}(f(t)\, e^{-pt})\, dt = -\frac{d}{dp}\int_0^\infty f(t)\, e^{-pt} = -\frac{d}{dp} F(p).$$

Следовательно,

$$tf(t) \doteqdot -\frac{d}{dp} F(p). \tag{7}$$

(Здесь использовано одно из свойств интегралов, зависящих от параметра.)

Следствие **1.** *Так как* $tf(t) \doteqdot -\frac{d}{dp} F(p)$, *то*

$$t^2 f(t) = t\,[tf(t)] \doteqdot -\frac{d}{dp}\left(-\frac{d}{dp} F(p)\right) = (-1)^2 \frac{d^2F(p)}{dp^2},$$

$$t^3 f(t) = t\,[t^2 f(t)] \doteqdot -\frac{d}{dp}\left((-1)^2 \frac{d^2F(p)}{dp^2}\right) = (-1)^3 \frac{d^3F(p)}{dp^3},$$

. . . . . .

$$t^n f(t) \doteqdot (-1)^n \frac{d^nF(p)}{dp^n}. \tag{8}$$

Следствие **2.** *Известно, что* $\sigma(t) \doteqdot \frac{1}{p}$ *и*

$$t^n f(t) \doteqdot (-1)^n \frac{d^nF(p)}{dp^n}.$$

Поэтому

$$t = t \cdot \sigma(t) \doteqdot -\frac{d}{dp}\left(\frac{1}{p}\right) = -\frac{1}{p^2}(-1),$$

$$t^2 = t \cdot t \doteqdot -\frac{d}{dt}\left(\frac{1}{p^2}\right) = (-1)^2 \frac{2!}{p^3}(-1)^2,$$

$$t^3 = t \cdot t^2 \doteqdot -\frac{d}{dp}\left((-1)^2 \frac{2!}{p^3}\right) = (-1)^3 \frac{3!}{p^4}(-1)^3,$$

. . . . . .

$$t^n \doteqdot \frac{n!}{p^{n+1}}. \qquad (9)$$

Из этого операционного соотношения, используя свойство линейности, можно получить и такое соответствие:

$$\frac{1}{p^{n+1}} \doteqdot \frac{t^n}{n!}.$$

Пример. Найти изображение функции $t \sin \omega t$.
Р е ш е н и е. Известно, что

$$\sin \omega t \doteqdot \frac{\omega}{p^2 + \omega^2}.$$

Поэтому

$$t \sin \omega t \doteqdot -\frac{d}{dp}\left(\frac{\omega}{p^2 + \omega^2}\right) = \frac{2\omega p}{(p^2 + \omega^2)^2}.$$

Теорема об интегрировании изображения (о делении оригинала на аргумент). *Если* $\frac{f(t)}{t} \in D$ *и* $f(t) \doteqdot F(p)$, *то*

$$\frac{f(t)}{t} \doteqdot \int_p^\infty F(p)\,dp.$$

Д о к а з а т е л ь с т в о.

Обозначим $\frac{f(t)}{t} = \varphi(t)$ и $\varphi(t) \doteqdot \Phi(p)$.

Очевидно, $f(t) = t\varphi(t)$, и по предыдущей теореме

$$t\varphi(t) \doteqdot -\frac{d}{dp}\Phi(p).$$

По теореме единственности

$$F(p) = -\frac{d}{dp}\Phi(p).$$

Поэтому

$$\Phi(p) = -\int_0^p F(p)\,dp + C.$$

Произвольная постоянная интегрирования определяется из условия $\begin{cases}\Phi(p) \to 0,\\ p \to \infty,\end{cases}$ а именно:

$$\lim_{p\to\infty} \Phi(p) = \lim_{p\to\infty}\left(C - \int_0^p F(p)\,dp\right) = C - \int_0^\infty F(p)\,dp,$$

$$0 = C - \int_0^\infty F(p)\,dp, \quad C = \int_0^\infty F(p)\,dp.$$

Таким образом,

$$\Phi(p) = C - \int_0^p F(p)\,dp = \int_0^\infty F(p)\,dp - \int_0^p F(p)\,dp,$$

$$\Phi(p) = \int_p^\infty F(p)\,dp \doteqdot \frac{f(t)}{t}. \qquad (10)$$

Пример. Найти изображение функции $\frac{\operatorname{sh} t}{t}$.

Р е ш е н и е. Известно, что

$$\operatorname{sh} t \doteqdot \frac{1}{p^2-1},$$

поэтому

$$\frac{\operatorname{sh} t}{t} \doteqdot \int_p^\infty \frac{dp}{p^2-1} = \frac{1}{2}\ln\frac{p-1}{p+1}\Big|_p^\infty = -\frac{1}{2}\ln\frac{p-1}{p+1}.$$

Итак,

$$\frac{\operatorname{sh} t}{t} \doteqdot \frac{1}{2}\ln\frac{p+1}{p-1}.$$

Теорема подобия (об изменении масштаба). *Если* $f(t) \in D$ *и* $f(t) \doteqdot F(p)$, *то при* $\alpha > 0$ *всегда* $f\left(\frac{t}{\alpha}\right) \doteqdot \alpha F(p\alpha)$.

Д о к а з а т е л ь с т в о. Прежде всего отметим, что из $f(t) \in D$ следует

$$f\left(\frac{t}{\alpha}\right) \in D \quad (\alpha > 0).$$

Далее,

$$f\left(\frac{t}{\alpha}\right) = \int_0^\infty f\left(\frac{t}{\alpha}\right) e^{-pt}\,dt = \left|\begin{array}{l}\frac{t}{\alpha} = z,\ t = \alpha z,\ dt = \alpha dz,\\ \text{при } t = 0,\ z = 0,\\ \qquad t = \infty,\ z = \infty\end{array}\right| =$$

$$= \int_0^\infty f(z)\, e^{-p\alpha z}\alpha\,dz = \alpha F(p\alpha). \qquad (11)$$

Следовательно,

$$f\left(\frac{t}{\alpha}\right) \doteqdot \alpha F(p\alpha).$$

Пример. Известно, что $e^t \doteqdot \frac{1}{p-1}$. Найти изображение функции $e^{\alpha t}$.

Р е ш е н и е. Имеем

$$e^{\alpha t} = e^{\frac{t}{1/\alpha}} \doteqdot \frac{1}{\alpha}\,\frac{1}{p\cdot\frac{1}{\alpha}-1} = \frac{1}{p-\alpha}.$$

Теорема запаздывания. *Если* $f(t) \in D$, $f(t) \doteqdot F(p)$ *и* $\alpha > 0$, *то*

$$f(t-\alpha) \doteqdot e^{-\alpha p} F(p).$$

Доказательство. Так как $f(t) \equiv 0$ при $t < 0$ ($f(t) \in D$), то $f(t-\alpha) \equiv 0$ при $t < \alpha$ и $f(t-\alpha) \in D$.

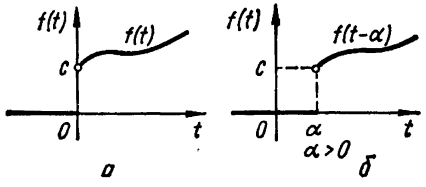

Рис. 42

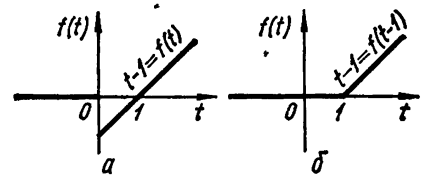

Рис. 43

Поэтому, производя замену переменной интегрирования по формуле $z = t - \alpha$, получаем

$$f(t-\alpha) \doteqdot \int_0^\infty f(t-\alpha)\, e^{-pt}\, dt = \int_\alpha^\infty f(t-\alpha)\, e^{-pt}\, dt =$$

$$= \left| \begin{array}{l} z = t-\alpha,\ dz = dt, \\ \{\alpha < t < \infty, \\ \{0 < z < \infty \end{array} \right| = \int_0^\infty f(z)\, e^{-p(z+\alpha)}\, dz =$$

$$= e^{-p\alpha} \int_0^\infty f(z)\, e^{-pz}\, dz = e^{-\alpha p} F(p).$$

Таким образом,

$$f(t-\alpha) \doteqdot e^{-\alpha p} F(p). \qquad (12)$$

Геометрический смысл запаздывания таков: функции $f(t)$ и $f(t-\alpha)$ имеют одинаковые графики (они описывают один и тот же процесс), но график функции $f(t-\alpha)$ сдвинут на α единиц вправо (процесс для этой функции начинается на α единиц позже, чем для функции $f(t)$), что и показано на рис. 42, *а*, *б*.

Пример. Найти изображение функции $f(t) = t - 1$.

Решение. Зная, что $t \doteqdot \frac{1}{p^2}$, $1 \doteqdot \frac{1}{p}$, и используя свойство линейности, находим

$$t - 1 \doteqdot \frac{1}{p^2} - \frac{1}{p}.$$

При отыскании изображения функции $t - 1$ возможен и другой подход.

Если считать, что задана функция $f_1(t) = t$, то $t - 1$ предстает как $f_1(t-1)$, т. е. $f_1(t-1) = t - 1$. Тогда изображение разности $t - 1$ можно получить с помощью теоремы запаздывания. Имеем

$$f_1(t) = t \doteqdot \frac{1}{p^2},$$

$$t - 1 = f_1(t-1) \doteqdot e^{-p} \frac{1}{p^2}.$$

Получились два различных результата, что на первый взгляд противоречит теореме единственности. Однако здесь никакого противоречия нет, ибо в рассмотренных случаях были найдены изображения различных функций (рис. 43, *а*, *б*).

Теорема упреждения. *Если* $f(t) \in D$, $f(t) \doteqdot F(p)$ *и* $\alpha > 0$, *то* $f(t+\alpha) \doteqdot e^{p\alpha}\left(F(p) - \int\limits_0^\alpha f(t)\, e^{-pt}\, dt\right)$.

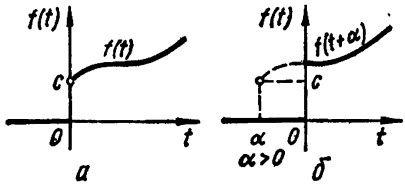

Рис. 44

Доказательство. Учитывая, что $f(t)|_{t<0} \equiv 0$, и применяя замену переменной интегрирования, получаем

$$f(t+\alpha) \doteqdot \int\limits_0^\infty f(t+\alpha)\, e^{-pt}\, dt =$$

$$= \left| \begin{array}{c} t+\alpha = z,\ t = z-\alpha,\ dt = dz, \\ 0 < t < \infty, \\ \alpha < z < \infty \end{array} \right| =$$

$$= \int\limits_\alpha^\infty f(z)\, e^{-p(z-\alpha)}\, dz = e^{p\alpha}\left(\int\limits_0^\infty f(z)\, e^{-pz}\, dz - \int\limits_0^\alpha f(z)\, e^{-pz}\, dz\right) =$$

$$= e^{p\alpha}\left(F(p) - \int\limits_0^\alpha f(t)\, e^{-pt}\, dt\right). \qquad (13)$$

Теорема доказана.

Функция $f(t+\alpha)$, для которой здесь получено изображение, представляет собой как оригинал функцию

$$f_1(t) = f(t+\alpha)\,\sigma(t) = \begin{cases} f(t+\alpha) & \text{при } t > 0, \\ 0 & \text{при } t < 0 \end{cases}$$

(на основании замечания 1, с. 104).

График функции $f_1(t)$ — это график функции $f(t)$, смещенный влево на α при условии, что часть, оказавшаяся левее начала координат, заменена отрезком оси абсцисс (рис. 44, *а*, *б*).

Теорема смещения. *Если* $f(t) \in D$, $f(t) \doteqdot F(p)$ *и* α — *произвольная постоянная, то*

$$e^{-\alpha t} f(t) \doteqdot F(p+\alpha).$$

Доказательство. Очевидно, $e^{-\alpha t} f(t) \in D$, поэтому

$$e^{-\alpha t} f(t) \doteqdot \int\limits_0^\infty e^{-\alpha t} f(t)\, e^{-pt}\, dt = \int\limits_0^\infty f(t)\, e^{-(p+\alpha)t}\, dt = |\, p+\alpha = q\,| =$$

$$= \int\limits_0^\infty f(t)\, e^{-qt}\, dt = F(q) = F(p+\alpha).$$

Таким образом,

$$e^{-\alpha t} f(t) \doteqdot F(p+\alpha). \tag{14}$$

Теорема умножения (теорема о свертке). *Если* $f_1(t) \in D$, $f_2(t) \in D$, $f_1(t) \doteqdot F_1(p)$, $f_2(t) \doteqdot F_2(p)$, *то*

$$\int_0^t f_1(\tau) \cdot f_2(t-\tau)\, d\tau \doteqdot F_1(p) \cdot F_2(p).$$

Интеграл $\int_0^t f_1(\tau) \cdot f_2(t-\tau)\, d\tau = \int_0^t f_1(t-\tau) \times f(\tau)\, d\tau$ называется свертко й функций $f_1(t)$ и $f_2(t)$ и обозначается $f_1 \ast f_2$.

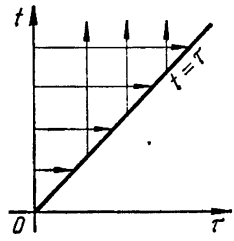

Рис. 45

Если $f_1(t)$ и $f_2(t)$ — оригиналы, то свертка их также является оригиналом.

Действительно, а) если числа $M_1 > 0$ и $S_1 > 0$ таковы, что

$$|f_1(t)| \leqslant M_1 e^{S_1 t}, \quad |f_2(t)| \leqslant M_1 e^{S_1 t},$$

то

$$\left| \int_0^t f_1(\tau) \cdot f_2(t-\tau)\, d\tau \right| \leqslant \int_0^t |f_1(\tau)| \cdot |f_2(t-\tau)|\, d\tau \leqslant$$

$$\leqslant M_1^2 \int_0^t e^{S_1 \tau} e^{S_1 (t-\tau)}\, d\tau = M_1^2 e^{S_1 t} \cdot t \leqslant M e^{S_0 t},$$

где $S_0 > S_1$ и M — достаточно велико;

б) при отрицательных аргументах функции $f_1(\tau)$ и $f_2(t-\tau)$ обращаются в нуль; свертка $f_1 \ast f_2$ также при $t < 0$ обращается в нуль;

в) можно показать, что благодаря непрерывности функций $f_1(t)$ и $f_2(t)$ их свертка $f_1 \ast f_2$ также непрерывна (как неопределенный интеграл от непрерывной функции — произведения двух непрерывных функций).

Таким образом, свертка двух оригиналов является оригиналом.

Найдем соответствующее изображение:

$$f_1 \ast f_2 \doteqdot \int_0^\infty e^{-pt}\, dt \int_0^t f_1(\tau) f_2(t-\tau)\, d\tau.$$

Меняя порядок интегрирования (область интегрирования представлена на рис. 45), получаем

$$f_1 \ast f_2 \doteqdot \int_0^\infty d\tau \int_\tau^\infty f_1(\tau) f_2(t-\tau) e^{-pt}\, dt =$$

$$= \int_0^\infty f_1(\tau)\, d\tau \int_\tau^\infty f_2(t-\tau) e^{-pt}\, dt = \left| \begin{array}{l} t-\tau = z, \ t = z+\tau, \\ dt = dz, \\ \begin{cases} \tau \leqslant t < \infty, \\ 0 \leqslant z < \infty \end{cases} \end{array} \right| =$$

$$= \int_0^\infty f_1(\tau)\, d\tau \int_0^\infty f_2(z) e^{-p(z+\tau)}\, dz =$$

$$= \int_0^\infty f_1(\tau)\, e^{-p\tau}\, d\tau \int_0^\infty f_2(z)\, e^{-pz}\, dz = F_1(p)\, F_2(p).$$

Следовательно,

$$f_1 \ast f_2 \doteqdot F_1(p)\, F_2(p). \tag{15}$$

Пример. Применяя теорему умножения, найти оригинал, если

$$F(p) = \frac{p}{(p^2+1)^2}.$$

Решение. Имеем

$$\frac{p}{(p^2+1)^2} = \frac{p}{p^2+1} \cdot \frac{1}{p^2+1}, \quad \frac{p}{p^2+1} \doteqdot \cos t, \quad \frac{1}{p^2+1} \doteqdot \sin t.$$

Поэтому, считая

$$F_1(p) = \frac{p}{p^2+1} \doteqdot \cos t = f_1(t),$$

$$F_2(p) = \frac{1}{p^2+1} \doteqdot \sin t = f_2(t),$$

находим

$$F_1(p) \cdot F_2(p) \doteqdot f_1(t) \ast f_2(t) = \int_0^t f_1(\tau)\, f_2(t-\tau)\, d\tau,$$

т. е.

$$\frac{p}{(p^2+1)^2} = \frac{p}{p^2+1} \cdot \frac{1}{p^2+1} \doteqdot \int_0^t \cos\tau \sin(t-\tau)\, d\tau =$$

$$= \frac{1}{2} \int_0^t (\sin(t-\tau-\tau) + \sin t)\, d\tau = \frac{1}{2}\, t \sin t.$$

Первая теорема разложения. *Если функция $F(p)$ представляет собой степенной ряд по отрицательным степеням p с ненулевым радиусом сходимости, т. е.*

$$F(p) = \sum_{n=0}^{\infty} \frac{a_n}{p^{n+1}}, \tag{16}$$

то оригиналом такой функции является степенной ряд по положительным степеням t — ряд

$$\sum_{n=0}^{\infty} a_n \frac{t^n}{n!}, \tag{17}$$

сходящийся при всех $t > 0$:

$$F(p) = \sum_{n=0}^{\infty} \frac{a_n}{p^{n+1}} \doteqdot \sum_{n=0}^{\infty} a_n \frac{t^n}{n!}. \qquad (18)$$

Примем эту теорему без доказательства. Отметим только, что формальное соответствие рядов (16) и (17) обеспечено, так как $\frac{a_n}{p^{n+1}} \doteqdot \frac{a_n}{n!} t^n$, а доказательство должно привести к следующему утверждению: *ряд* (17) *равномерно сходится при всех* $t > 0$, *и сумма его равна оригиналу* $f(t)$ *для функции* $F(p)$.

Вторая теорема разложения. *Если* $F(p) = \frac{P(p)}{Q(p)}$ — *рациональная правильная несократимая дробь, знаменатель* $Q(p)$ *которой имеет лишь простые корни* $\alpha_1, \alpha_2, \ldots, \alpha_n$, *то*

$$F(p) = \frac{P(p)}{Q(p)} \doteqdot \sum_{k=1}^{n} A_k e^{\alpha_k t},$$

где

$$A_k = \frac{P(\alpha_k)}{Q'(\alpha_k)}, \quad Q'(p) = \frac{dQ(p)}{dp}.$$

Доказательство. Прежде всего заметим, что требование правильности дроби в данной теореме обязательно, так как эта дробь — изображение, и должно быть выполнено условие $\lim\limits_{p \to \infty} F(p) = 0$.

Далее известно, что в случае простых корней знаменателя правильная рациональная дробь может быть разложена на простейшие следующим образом:

$$\frac{P(p)}{Q(p)} = \frac{A_1}{p - \alpha_1} + \frac{A_2}{p - \alpha_2} + \cdots + \frac{A_n}{p - \alpha_n}, \qquad (19)$$

где α_k $(k = 1, 2, 3, \ldots, n)$ — простые корни знаменателя (простые полюсы функции $F(p)$), а коэффициенты $A_k = \text{const}$ — числа, подлежащие определению. Для нахождения коэффициента A_1 домножим обе части равенства (19) на двучлен $(p - \alpha_1)$. Получим

$$\frac{P(p)(p - \alpha_1)}{Q(p)} =$$

$$= A_1 + (p - \alpha_1)\left(\frac{A_2}{p - \alpha_2} + \frac{A_3}{p - \alpha_3} + \cdots + \frac{A_n}{p - \alpha_n}\right). \qquad (20)$$

Введем обозначение $\frac{Q(p)}{p - \alpha_1} = Q_1(p)$, тогда предыдущее равенство (20) можно записать в виде

$$\frac{P(p)}{Q_1(p)} = A_1 + (p - \alpha_1)\left(\frac{A_2}{p - \alpha_2} + \frac{A_3}{p - \alpha_3} + \cdots + \frac{A_n}{p - \alpha_n}\right).$$

Полагая $p = \alpha_1$, найдем $A_1 = \frac{P(\alpha_1)}{Q_1(\alpha_1)}$. Легко проверить, что $Q_1(p)|_{p=\alpha_1} = \left(\frac{d}{dp} Q(p)\right)\Big|_{p=\alpha_1} = Q'(\alpha_1)$. Поэтому $A_1 = \frac{P(\alpha_1)}{Q'(\alpha_1)}$.

Подобно этому вычисляются и все остальные коэффициенты разложения:

$$A_k = \frac{P(\alpha_k)}{Q'(\alpha_k)} \qquad (k = 1, 2, 3, \ldots, n). \tag{21}$$

Используя теперь теорему смещения $e^{\alpha t} f(t) \doteqdot F(p - \alpha)$ (при $f(t) \doteqdot F(p)$) и учитывая, что $1 = \sigma(t) \doteqdot \frac{1}{p}$, приходим к выводу:

$$\frac{A_k}{p - \alpha_k} \doteqdot A_k e^{\alpha_k t}.$$

Итак,

$$F(p) = \frac{P(p)}{Q(p)} \doteqdot \sum_{k=1}^{n} A_k e^{\alpha_k t}. \tag{22}$$

Рассмотрим примеры применения теорем разложения.

Примеры

1. Найти оригинал $f(t)$ функции $F(p) = \ln\left(1 - \frac{1}{p}\right)$.

Р е ш е н и е. Разложим $F(p)$ в степенной ряд:

$$F(p) = \ln\left(1 - \frac{1}{p}\right) = -\frac{1}{p} - \frac{1}{2p^2} - \ldots - \frac{1}{np^n} - \ldots = -\sum_{n=1}^{\infty} \frac{1}{np^n}.$$

Этот ряд сходится при $p > 1$, что можно проверить, применяя, например, признак д'Аламбера. В соответствии с первой теоремой разложения получаем

$$F(p) = -\sum_{n=1}^{\infty} \frac{1}{np^n} \doteqdot -\sum_{n=1}^{\infty} \frac{1}{n} \cdot \frac{t^{n-1}}{(n-1)!} = -\sum_{n=1}^{\infty} \frac{t^{n-1}}{n!} = \frac{1 - e^t}{t}.$$

2. Найти оригинал $f(t)$ функции $F(p) = \frac{p - 1}{(p + 1)(p^2 + 4)}$.

Р е ш е н и е. $F(p)$ — правильная рациональная несократимая дробь, причем

$$P(p) = p - 1,$$
$$Q(p) = (p + 1)(p^2 + 4) = p^3 + p^2 + 4p + 4.$$

Корни знаменателя простые: $\alpha_1 = -1$, $\alpha_2 = 2i$, $\alpha_3 = -2i$; поэтому можно применить вторую теорему разложения. Очевидно, $Q'(p) = 3p^2 + 2p + 4$. Все дальнейшие действия удобно свести в таблицу:

Таблица 1

	$\alpha_1 = -1$	$\alpha_2 = 2i$	$\alpha_3 = -2i$
$P(\alpha_k) = p - 1\,\big\vert_{p=\alpha_k}$	-2	$-1 + 2i$	$-1 - 2i$
$Q'(\alpha_k) = 3p^2 + 2p + 4\,\big\vert_{p=\alpha_k}$	5	$-8 + 4i$	$-8 - 4i$
$A_k = \frac{P(\alpha_k)}{Q'(\alpha_k)}$	$-\frac{2}{5}$	$\frac{1 - 2i}{8 - 4i} = \frac{4 - 3i}{20}$	$\frac{1 + 2i}{8 + 4i} = \frac{4 + 3i}{20}$

В итоге получаем:

$$f(t) = -\frac{2}{5}e^{-t} + \frac{4-3i}{20}e^{2it} + \frac{4+3i}{20}e^{-2it} =$$

$$= -\frac{2}{5}e^{-t} + \frac{2}{5}\cos 2t + \frac{3}{10}\sin 2t.$$

Изображение периодических функций. Пусть $f(t)$ — периодическая функция с периодом T, т. е. $f(t - T) = f(t)$, тогда она может быть представлена в виде сходящегося ряда

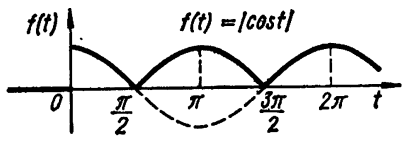

Рис. 46

$$f(t) = \sum_{k=0}^{\infty} f_0(t - kT), \tag{23}$$

где

$$f_0(t) = \begin{cases} f(t) & \text{при } 0 < t < T, \\ 0 & \text{при } t < 0 \text{ и } t > T. \end{cases} \tag{24}$$

Если $f(t) \in D$ и $f_0(t) \doteqdot F_0(p)$, то на основании теоремы запаздывания получаем

$$f_0(t - T) \doteqdot e^{-pT}F_0(p),$$
$$f_0(t - 2T) \doteqdot e^{-2pT}F_0(p),$$
$$\cdots\cdots\cdots\cdots$$
$$f_0(t - kT) \doteqdot e^{-kpT}F_0(p),$$
$$\cdots\cdots\cdots\cdots$$

Поэтому при достаточно больших p ($\operatorname{Re} p > S_0$)

$$\sum_{k=0}^{\infty} f_0(t - kT) \doteqdot \sum_{k=0}^{\infty} e^{-kpT}F_0(p) = F_0(p)\sum_{k=0}^{\infty} e^{-kpT} = F_0(p)\frac{1}{1 - e^{-pT}}.$$

Итак,

$$f(t) \doteqdot \frac{F_0(p)}{1 - e^{-pT}}, \tag{25}$$

где $F_0(p)$ — изображение функции $f(t)$ на первом (начальном) периоде.

Пример. Найти изображение функции $f(t) = |\cos t|$, $0 < t < \infty$ (рис. 46).
Решение. Здесь

$$f_0(t) = \begin{cases} |\cos t| & \text{при } 0 < t < \pi, \\ 0 & \text{при } t < 0 \text{ и } t > \pi. \end{cases}$$

При этом

$$f(t) = f(t - \pi k), \quad f(t) = \sum_{k=0}^{\infty} f_0(t - \pi k).$$

Имеем

$$f_0(t) \doteqdot F_0(p) = \int_0^{\pi} |\cos t| e^{-pt} dt = \int_0^{\pi/2} \cos t e^{-pt} dt - \int_{\pi/2}^{\pi} \cos t e^{-pt} dt =$$

$$= \frac{1}{p^2+1}\left(2e^{-p\frac{\pi}{2}} + p(1-e^{-p\pi})\right).$$

Поэтому

$$f(t) = |\cos t| \doteqdot \frac{F_0(p)}{1-e^{-p\pi}} = \frac{2e^{-p\frac{\pi}{2}} + p(1-e^{-p\pi})}{(p^2+1)(1-e^{-p\pi})}.$$

Интеграл Дюамеля. *Если* $f(t) \ast \varphi(t) = \int_0^t f(t-\tau)\varphi(\tau)\,d\tau \doteqdot F(p)\Phi(p)$, *где* $f(t) \doteqdot F(p)$ *и* $\varphi(t) \doteqdot \Phi(p)$, *то*

$$\int_0^t f(\tau)\varphi_t'(t-\tau)\,d\tau + f(t)\varphi(0) \doteqdot pF(p)\Phi(p),$$

или

$$\int_0^t f_t'(t-\tau)\varphi(\tau)\,d\tau + f(0)\varphi(t) \doteqdot pF(p)\Phi(p).$$

Доказательство. Из курса математического анализа известно, что

$$\frac{d}{dt}\int_{\alpha(t)}^{\beta(t)} f(x,t)\,dx = \int_{\alpha(t)}^{\beta(t)} \frac{\partial f(x,t)}{\partial t}\,dx + f(\beta(t),t)\frac{d\beta}{dt} - f(\alpha(t),t)\frac{d\alpha}{dt}$$

(дифференцирование по параметру интеграла, зависящего от параметра). Поэтому для свертки двух функций имеем

$$\frac{d}{dt}(f(t) \ast \varphi(t)) = \frac{d}{dt}\int_0^t f(\tau)\varphi(t-\tau)\,d\tau = \int_0^t f(\tau)\varphi_t'(t-\tau)\,d\tau +$$

$$+ f(t)\varphi(t-t) - 0 = \int_0^t f(\tau)\varphi_t'(t-\tau)\,d\tau + f(t)\varphi(0).$$

С другой стороны, так как $f(t) \ast \varphi(t) \doteqdot F(p)\Phi(p)$, то $\frac{d}{dt}(f(t) \ast \varphi(t)) \doteqdot p(F(p)\Phi(p)) - (f(t) \ast \varphi(t))|_{t=0} = pF(p)\Phi(p)$, поскольку

$$(f(t) \ast \varphi(t))|_{t=0} = \int_0^0 f(\tau)\varphi(t-\tau)\,d\tau = 0.$$

Из этого следует, что

$$\int_0^t f(\tau)\,\varphi_t'(t-\tau)\,d\tau + f(t)\,\varphi(0) \doteqdot pF(p)\,\Phi(p), \qquad (26)$$

и, подобно последнему,

$$\int_0^t f_t'(t-\tau)\,\varphi(\tau)\,d\tau + f(0)\,\varphi(t) \doteqdot pF(p)\,\Phi(p). \qquad (27)$$

Интегралы, стоящие в формулах (26) и (27), называют интегралами Дюамеля, или интегралами наложения.

Пример. Найти оригинал $g(t)$ изображения $G(p) = \dfrac{p^2}{(p^2+4)^2}$.

Решение. Так как

$$G(p) = p\,\frac{p}{p^2+4}\cdot\frac{1}{p^2+4}, \quad \frac{p}{p^2+4} \doteqdot \cos 2t, \quad \frac{1}{p^2+4} \doteqdot \frac{1}{2}\sin 2t$$

и

$$pF(p)\cdot\Phi(p) \doteqdot \int_0^t f_t'(t-\tau)\cdot\varphi(\tau)\,d\tau + f(0)\cdot\varphi(t),$$

где

$$F(p) \doteqdot f(t), \quad \Phi(p) \doteqdot \varphi(t),$$

то, считая

$$\cos 2t = \varphi(t) \doteqdot \Phi(p),$$

$$\sin 2t = f(t) \doteqdot F(p),$$

найдем

$$G(p) = pF(p)\cdot\Phi(p) \doteqdot \frac{1}{2}\left[\int_0^t 2\cos(2t-2\tau)\cos 2\tau\,d\tau\right] =$$

$$= \frac{1}{2}\int_0^t [\cos(2t-2\tau-2\tau) + \cos(2t-2\tau+2\tau)]\,d\tau = \frac{1}{4}\sin 2t + \frac{1}{2}\,t\cos 2t.$$

Изображение дельта-функции $\delta(t)$. Определение этой функции было дано в главе III.

По свойствам дельта-функции для всякой непрерывной функции $f(t)$ справедливо равенство

$$\int_a^b f(t)\,\delta(t-t_0)\,dt = f(t_0).$$

Поэтому, непосредственно применяя интеграл Лапласа, получаем

$$\int_0^\infty \delta(t)\,e^{-pt}dt = \int_0^\infty e^{-pt}\delta(t-0)\,dt = e^{-p\cdot 0} = 1,$$

т. е.

$$\delta(t) \doteqdot 1. \tag{28}$$

Подчеркнем, что это изображение получается в условиях, когда сама функция $\delta(t)$ неограниченно возрастает. Оказывается, что как сама функция $\delta(t)$ не принадлежит классу оригиналов в принятых нами условиях ($\delta(t)$ имеет бесконечный разрыв, хотя и является предельной функцией для последовательности непрерывных функций), так и ее изображение обладает не всеми свойствами изображений (оно не стремится к нулю при $p \to \infty$).

Изображение кусочно-непрерывных и кусочно-гладких функций. а) Изображение единичной функции

$$\sigma(t) = \begin{cases} 0 & \text{при } t < 0, \\ 1 & \text{при } t \geqslant 0 \end{cases}$$

было получено выше (см. § 17).

$$\sigma(t) \doteqdot \frac{1}{p}. \tag{29}$$

Может встретиться единичная функция с «запаздыванием» (рис. 47).

$$\sigma(t-\alpha) = \begin{cases} 0 & \text{при } t < \alpha, \\ 1 & \text{при } t > \alpha \quad (\alpha > 0). \end{cases}$$

По теореме запаздывания имеем

$$\sigma(t-\alpha) \doteqdot \frac{e^{-\alpha p}}{p}. \tag{30}$$

б) Изображение функции, график которой состоит из отрезка оси Ot и гладкой кривой (рис. 48),

$$f(t) = \begin{cases} 0 & \text{при } t < \alpha, \\ \varphi(t) & \text{при } t > \alpha \end{cases}$$

может быть получено непосредственно по интегралу Лапласа:

$$f(t) \doteqdot F(p) = \int_0^\infty f(t)\, e^{-pt} dt = \int_\alpha^\infty \varphi(t)\, e^{-pt} dt = \left| \begin{array}{l} z = t - \alpha,\ dt = dz, \\ t = z + \alpha, \\ \begin{cases} \alpha < t < \infty, \\ 0 < z < \infty \end{cases} \end{array} \right| =$$

$$= \int_0^\infty \varphi(z+\alpha)\, e^{-p(z+\alpha)} dz = e^{-p\alpha} \int_0^\infty \varphi(z+\alpha)\, e^{-pz} dz = e^{-\alpha p} \Phi(p, \alpha),$$

где $\Phi(p, \alpha)$ — изображение функции $\varphi(z+\alpha)$.

То же изображение может быть найдено (и это удобнее в практическом применении) с помощью единичной функции $\sigma(t)$ в результате использования теоремы запаздывания.

Для этого функцию $\varphi(t)$ представляют в окрестности точки $z = \alpha$ в виде суммы ряда по произвольно выбранной системе функ-

ций $\varphi_k(t)$, изображения которых известны. (Такие функции называются координатными функциями. Ряд по выбранной системе функций может быть конечным или бесконечным — это зависит от самой функции $\varphi(t)$ и координатных функций $\varphi_k(t)$; последние лучше выбирать так, чтобы получился конечный ряд, тогда отпадает необходимость исследовать сходимость ряда изображений).

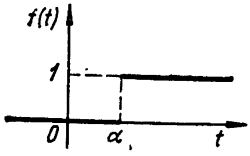

Рис. 47

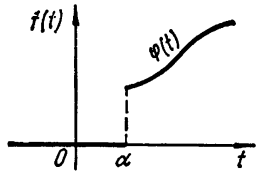

Рис. 48

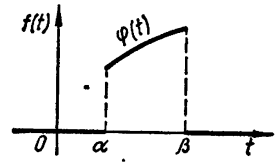

Рис. 49

Затем представляют функцию $f(t)$, учитывая запаздывание, и находят изображение этой функции как сумму ряда изображений функций $\varphi_k(t-\alpha)$ и $\sigma(t-\alpha)$. В аналитической записи это выглядит так.

Пусть $\varphi(t) = \sum_{k=1}^{n} A_k \varphi_k(t-\alpha)$, причем $\varphi_k(t)\,\sigma(t) \doteqdot \Phi_k(p)$, тогда

$$f(t) = \left(\sum_{k=1}^{n} A_k \varphi_k(t-\alpha)\right)\sigma(t-\alpha) \doteqdot e^{-p\alpha} \sum_{k=1}^{n} \Phi_k(p, \alpha).$$

Если $\sum_{k=1}^{n} \Phi_k(p, \alpha) = \Phi(p, \alpha)$, то окончательно

$$f(t) \doteqdot e^{-\alpha p}\Phi(p, \alpha). \tag{31}$$

в) Изображение функции, график которой состоит из изолированной гладкой дуги и двух отрезков оси абсцисс (рис. 49)

$$f(t) = \begin{cases} 0 & \text{при } 0 < t < \alpha, \\ \varphi(t) & \text{при } \alpha < t < \beta, \\ 0 & \text{при } t > \beta \end{cases}$$

может быть найдено непосредственно с помощью интеграла Лапласа:

$$F(p) = \int_0^{\infty} f(t)\,e^{-pt}dt = \int_{\alpha}^{\beta} \varphi(t)\,e^{-pt}dt = \int_{\alpha}^{\infty} \varphi(t)\,e^{-pt}dt - \int_{\beta}^{\infty} \varphi(t)\,e^{-pt}dt,$$

откуда

$$F(p) = e^{-\alpha p}\Phi(p, \alpha) - e^{-\beta p}\Phi(p, \beta). \tag{32}$$

Это изображение является разностью изображений функции $f(t)$ с запаздыванием до значения $t = \alpha$ и с запаздыванием до значения $t = \beta$.

Но здесь удобно использовать рассмотренный выше способ представления «запаздывающей» функции, а именно, выбрав систему координатных функций, надо найти разложения функции $\varphi(t)$ по этой системе в точках $t = \alpha$ и $t = \beta$

$$\varphi(t) = \sum_{k=1}^{n} A_k \varphi_k (t - \alpha) = \sum_{k=1}^{n} B_k \varphi_k (t - \beta);$$

затем представить с помощью этих разложений $f(t)$:

$$f(t) = \left(\sum_{k=1}^{n} A_k \varphi_k (t - \alpha) \right) \sigma(t - \alpha) - \left(\sum_{k=1}^{n} B_k \varphi_k (t - \beta) \right) \sigma(t - \beta);$$

и, наконец, найти изображение функции $f(t)$, как это показано в предыдущем пункте.

Примеры

1. Найти изображение функции (рис. 50)

$$f(t) = \begin{cases} h & \text{при } 1 < t < 3, \\ 0 & \text{при } t < 1 \text{ и } t > 3. \end{cases}$$

Решение. Имеем

$$f(t) = h\sigma(t - 1) - h\sigma(t - 3) \doteqdot \frac{h}{p}(e^{-p} - e^{-3p}).$$

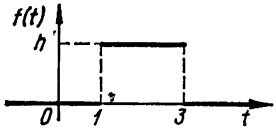

Рис. 50

2. Найти изображение функции (рис. 51)

$$f(t) = \begin{cases} 3t + 4 & \text{при } 2 < t < 5, \\ 0 & \text{при } t < 2 \text{ и } t > 5. \end{cases}$$

Решение. Выберем в качестве координатных функций степенные функции. Тогда $3t + 4 = 3(t - 2) + 10 = 3(t - 5) + 19$, $f(t) = (3(t - 2) + 10) \times \sigma(t - 2) - (3(t - 5) + 19)\sigma(t - 5)$.

Следовательно,

$$f(t) \doteqdot F(p) = \left(\frac{3}{p^2} + \frac{10}{p} \right) e^{-2p} - \left(\frac{3}{p^2} + \frac{19}{p} \right) e^{-5p}.$$

3. Найти изображение функции (рис. 52)

$$f(t) = \begin{cases} 0 & \text{при } t < 1, \\ t^2 & \text{при } 1 < t < 2, \\ 0 & \text{при } t > 2. \end{cases}$$

Решение. В качестве координатных функций выберем степенные функции. Тогда

$$f(t) = ((t - 1)^2 + 2(t - 1) + 1)\sigma(t - 1) - ((t - 2)^2 + 4(t - 2) + 4)\sigma(t - 2),$$

а потому

$$f(t) \doteqdot F(p) = \left(\frac{2}{p^3} + \frac{2}{p^2} + \frac{1}{p} \right) e^{-p} - \left(\frac{2}{p^3} + \frac{4}{p^2} + \frac{4}{p} \right) e^{-2p}.$$

4. Найти изображение функции (рис. 53)

$$f(t) = \begin{cases} 0 & \text{при } t < 2, \\ \sin t & \text{при } 2 < t < 3, \\ 0 & \text{при } t < 3. \end{cases}$$

Р е ш е н и е. В качестве координатных функций выберем синусы и косинусы. Известно, что

$$\sin t = \sin((t-2)+2) = \cos 2 \sin(t-2) + \sin 2 \cos(t-2)$$

и

$$\sin t = \sin((t-3)+3) = \cos 3 \sin(t-3) + \sin 3 \cos(t-3).$$

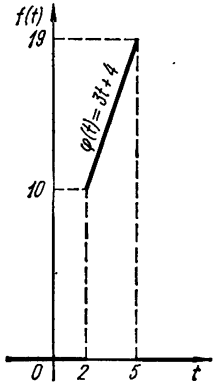

Рис. 51

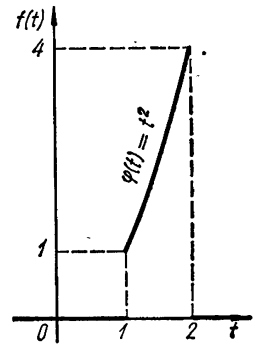

Рис. 52

Поэтому

$$f(t) = (\cos 2 \sin(t-2) + \sin 2 \cos(t-2))\,\sigma(t-2) -$$
$$-(\cos 3 \sin(t-3) + \sin 3 \cos(t-3))\,\sigma(t-3).$$

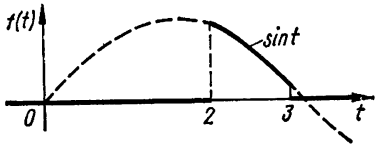

Рис. 53

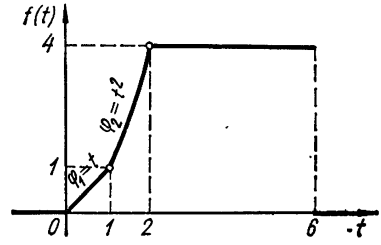

Рис. 54

Это дает

$$f(t) \doteqdot F(p) = \left(\frac{1}{p^2+1}\cos 2 + \frac{p}{p^2+1}\sin 2\right)e^{-2p} -$$
$$-\left(\frac{1}{p^2+1}\cos 3 + \frac{p}{p^2+1}\sin 3\right)e^{-3p}.$$

Замечание. Для получения изображения произвольной кусочно-непрерывной или кусочно-гладкой функции эту функцию расчленяют на непрерывные гладкие участки. Для каждого из этих участков находят соответствующее изображение, как это показано для простейшей кусочно-непрерывной функции, а затем полученные результаты суммируют.

5. Найти изображение функции (рис. 54)

$$f(t) = \begin{cases} t & \text{при } 0 \leqslant t \leqslant 1, \\ t^2 & \text{при } 1 \leqslant t \leqslant 2, \\ 4 & \text{при } 2 \leqslant t \leqslant 6. \end{cases}$$

Р е ш е н и е. Здесь имеются три участка непрерывности и гладкости функции. В каждом участке функция может быть записана с помощью единичной функции при обязательном применении теоремы запаздывания. Чтобы записать всю функцию, необходимо сложить все полученные представления. Тогда

$$f(t) = (t\sigma(t) - (t-1)\sigma(t-1)) + (((t-1)^2 + 2(t-1) + 1)\sigma(t-1) - $$
$$- ((t-2)^2 + 4(t-2) + 4)\sigma(t-2)) + (4\sigma(t-2) - 4\sigma(t-6)).$$

Поэтому

$$f(t) \doteqdot F(p) = \left(\frac{1}{p^2} - \frac{e^{-p}}{p^2}\right) + \left(\frac{2}{p^3} + \frac{2}{p^2} + \frac{1}{p}\right)e^{-p} -$$
$$- \left(\frac{2}{p^3} + \frac{4}{p^2} + \frac{4}{p}\right)e^{-2p} + \frac{4}{p}e^{-3p} - \frac{4}{p}e^{-6p}.$$

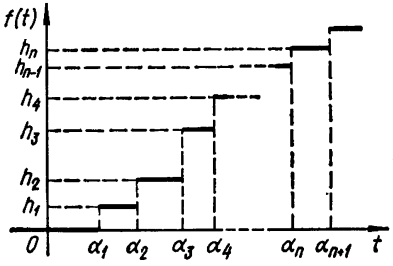

Рис. 55

Изображение ступенчатой функции. Ступенчатая функция — это функция вида (рис. 55)

$$f(t) = \begin{cases} 0 & \text{при } t < \alpha_1, \\ h_1 & \text{при } \alpha_1 < t < \alpha_2, \\ h_2 & \text{при } \alpha_2 < t < \alpha_3, \\ \cdots\cdots\cdots\cdots \\ h_n & \text{при } \alpha_n < t < \alpha_{n+1}. \end{cases}$$

Очевидно,

$$f(t) = (h_1\sigma(t-\alpha_1) - h_1\sigma(t-\alpha_2)) + (h_2\sigma(t-\alpha_2) - h_2\sigma(t-\alpha_3)) +$$
$$+ (h_3\sigma(t-\alpha_3) - h_3\sigma(t-\alpha_4)) + \cdots + (h_n\sigma(t-\alpha_n) -$$
$$- h_n\sigma(t-\alpha_{n+1})) = \sum_{k=1}^{n} h_k(\sigma(t-\alpha_k) - \sigma(t-\alpha_{k+1})) \doteqdot$$
$$\doteqdot \sum_{k=1}^{n} \frac{h_k}{p}(e^{-\alpha_k p} - e^{-\alpha_{k+1}p}).$$

Если принять $f(t) \doteqdot F(p)$ и обозначить $\alpha_{k+1} - \alpha_k = \Delta_k$, то

$$F(p) = \sum_{k=1}^{n} \frac{h_k}{p} e^{-\alpha_k p}(1 - e^{(\alpha_k - \alpha_{k+1})p}) = \frac{1}{p}\sum_{k=1}^{n} h_k e^{-\alpha_k p}(1 - e^{-\Delta_k p}).$$

§ 19. НЕКОТОРЫЕ ПРИЛОЖЕНИЯ ОПЕРАЦИОННОГО ИСЧИСЛЕНИЯ

В большинстве случаев применение операционного исчисления к решению конкретных примеров и задач укладывается в следующую общую схему.

Пусть требуется найти некоторый результат $\varphi(t)$, для получения которого (без использования операторного метода) надо над заданной функцией $f(t)$ выполнить какую-то операцию A.

Применяя операционное исчисление, сначала переходят от оригинала $f(t)$ к его изображению $F(p)$; затем над этим изображением

выполняют операцию B, соответствующую в операционном исчислении операции A (например, делят изображение на p вместо интегрирования функции $f(t)$), и получают промежуточный результат — изображение $\Phi(p)$; наконец, переходят от изображения $\Phi(p)$ к искомому оригиналу $\varphi(t)$.

Таким образом, общая схема решения задачи удлиняется. Однако здесь получается значительный выигрыш как в средствах вычисления, так и во времени. В частности, везде дифференцирование заменяется умножением на p и вычитанием, а интегрирование — делением на p. Применение начальных условий при решении дифференциальных уравнений ограничивается простым введением этих условий в составе изображений производных.

Этот выигрыш достигается путем широкого использования основных теорем операционного исчисления, свойств изображений и известных «табличных изображений», наиболее часто встречающихся функций.

Решение обыкновенных линейных дифференциальных уравнений с постоянными коэффициентами и систем таких уравнений. Решение указанных уравнений выполняется по схеме, изложенной выше. Покажем это на примере.

Пример. Найти частное решение дифференциального уравнения

$$y'' + 5y' + 6y = e^t \cos t,$$

удовлетворяющее начальным условиям $y(0) = 0$, $y'(0) = 1$.

Р е ш е н и е. 1) Находим «изображение уравнения» — уравнение в изображениях. Пусть $y(t) \doteqdot Y(p)$, тогда $y'(t) \doteqdot pY(p) - y(0) = pY(p)$,

$$y''(t) \doteqdot p^2Y(p) - py(0) - y'(0) = p^2Y(p) - 1.$$

В правой части заданного уравнения стоит выражение $e^t \cos t$.

Так как $\cos t \doteqdot \dfrac{p}{p^2+1}$, то $e^t \cos t \doteqdot \dfrac{p-1}{(p-1)^2+1}$.

Теперь на основании теоремы единственности получаем уравнение относительно изображения искомой функции:

$$p^2Y(p) - 1 + 5pY(p) + 6Y(p) = \frac{p-1}{(p-1)^2+1}.$$

2) Решаем получившееся алгебраическое (уже не дифференциальное!) уравнение относительно $Y(p)$.

Имеем

$$(p^2 + 5p + 6)\,Y(p) = \frac{p-1}{p^2-2p+2} + 1.$$

Отсюда

$$Y(p) = \frac{p^2 - p + 1}{(p^2+5p+6)(p^2-2p+2)}.$$

3) Определяем оригинал $y(t)$ — частное решение заданного дифференциального уравнения, удовлетворяющее заданным начальным условиям. Здесь знаменатель дроби

$$Q(p) = (p^2+5p+6)(p^2-2p+2) = p^4 + 3p^3 - 2p^2 - 2p + 12 =$$
$$= (p+3)(p+2)(p^2-2p+2)$$

имеет только простые корни $\alpha_1 = -3$, $\alpha_2 = -2$, $\alpha_3 = 1 + i$, $\alpha_4 = 1 - i$. Поэтому целесообразно применить вторую теорему разложения:

$$\frac{P(p)}{Q(p)} \doteqdot \sum_{k=1}^{4} A_k e^{\alpha_k t}, \quad A_k = \frac{P(\alpha_k)}{Q'(\alpha_k)}.$$

Имеем

$$P(p) = p^2 - p + 1,$$

$$Q(p) = p^4 + 3p^3 - 2p^2 - 2p + 12,$$

$$Q'(p) = 4p^3 + 9p^2 - 4p - 2.$$

Дальнейшую запись ведем в таблице:

Таблица 2

	$\alpha_1 = -3$	$\alpha_2 = -2$	$\alpha_3 = 1 + i$	$\alpha_4 = 1 - i$
$P(\alpha_k) = (p^2 - p + 1)\big\|_{p=\alpha_k}$	13	7	i	$-i$
$Q'(\alpha_k) = (4p^3 + 9p^2 - 4p - 2)\big\|_{p=\alpha_k}$	-17	10	$-14+22i$	$-14-22i$
$A_k = \frac{P(\alpha_k)}{Q'(\alpha_k)}$	$-\frac{13}{17}$	$\frac{7}{10}$	$\frac{-i}{14-22i}$	$\frac{i}{14+22i}$

Теперь получаем

$$y(t) = -\frac{13}{17} e^{-3t} + \frac{7}{10} e^{-2t} - \frac{i}{14-22i} e^{(1+i)t} + \frac{i}{14+22i} e^{(1-i)t}.$$

Выполняя тождественные преобразования, находим искомое решение

$$y(t) = -\frac{13}{17} e^{-3t} + \frac{7}{10} e^{-2t} + \frac{e^t}{170}(11\cos t + 7\sin t).$$

Так же выполняется решение систем обыкновенных линейных уравнений. Такие системы преобразуются к системам алгебраических линейных уравнений, решения которых являются изображениями решений исходных систем дифференциальных уравнений. Оригиналы этих решений находятся, как обычно.

Решение дифференциальных уравнений математической физики. Методы операционного исчисления применяются также и к решению дифференциальных уравнений с частными производными. Покажем, как это делается, на примере одномерного волнового уравнения.

Пусть требуется решить дифференциальное уравнение с частными производными:

$$\frac{\partial^2 u}{\partial t^2} - a^2 \frac{\partial^2 u}{\partial x^2} = f(x, t) \qquad (a = \text{const})$$

при начальных условиях

$$u(x, 0) = \varphi_1(x),$$

$$\left.\frac{\partial u}{\partial t}\right|_{t=0} = \varphi_2(x)$$

и граничных условиях

$$u|_{x=0} = \psi_1(t), \quad u|_{x=l} = \psi_2(t).$$

Предположим, что все упомянутые здесь функции, зависящие от t, принадлежат классу оригиналов и имеют соответствующие изображения:

$$u(x, t) \doteqdot \int_0^\infty u(x, t)\, e^{-pt} dt = U(x, p),$$

$$\frac{\partial u}{\partial t} \doteqdot pU(x, p) - u(x, 0) = pU(x, p) - \varphi_1(x),$$

$$\frac{\partial^2 u}{\partial t^2} \doteqdot p^2 U(x, p) - pu(x, 0) - \left.\frac{\partial u}{\partial t}\right|_{t=0} = p^2 U(x, p) - p\varphi_1(x) - \varphi_2(x),$$

$$f(x, t) \doteqdot \int_0^\infty f(x, t)\, e^{-pt} dt = F(x, p),$$

$$\frac{\partial u}{\partial x} \doteqdot \int_0^\infty \frac{\partial u(x, t)}{\partial x} e^{-pt} dt = \frac{\partial}{\partial x} \int_0^\infty u(x, t)\, e^{-pt} dt = \frac{\partial U(x, p)}{\partial x},$$

$$\frac{\partial^2 u}{\partial x^2} \doteqdot \int_0^\infty \frac{\partial^2 u(x, t)}{\partial x^2} e^{-pt} dt = \frac{\partial^2 U(x, p)}{\partial x^2},$$

$$\psi_1(t) \doteqdot \int_0^\infty \psi_1(t)\, e^{-pt} dt = \Psi_1(p),$$

$$\psi_2(t) \doteqdot \int_0^\infty \psi_2(t)\, e^{-pt} dt = \Psi_2(p).$$

Тогда легко получить «изображение уравнения» — уравнение в изображениях:

$$p^2 U(x, p) - p\varphi_1(x) - \varphi_2(x) - a^2 \frac{\partial^2 U(x, p)}{\partial x^2} = F(x, p)$$

при граничных условиях

$$U(x, p)|_{x=0} = \Psi_1(p), \quad U(x, p)|_{x=l} = \Psi_2(p).$$

Полученное уравнение — обыкновенное. Решая его, получим изображение искомой функции, по которому найдем и само решение исходного уравнения — функцию $u(x, t)$.

Решение операционными методами уравнений в конечных разностях

1. Разностные уравнения с непрерывным аргументом.

а) ***Определение разностей*** (более подробно о разностях см. главу IX). Пусть аргумент t получил приращение $\Delta t = \tau$. Тогда для функции $f(t)$ разностью первого порядка (или первой разностью) называется выражение

$$\Delta f(t) = f(t+\tau) - f(t). \tag{33}$$

И так как эта разность сама является функцией t, то, давая аргументу t новое значение $t+\alpha$, получим

$$\Delta f(t+\alpha) = f(t+\alpha+\tau) - f(t+\alpha),$$

а потому

$$\begin{aligned} \Delta f(t+\tau) &= f(t+2\tau) - f(t+\tau), \\ \Delta f(t+2\tau) &= f(t+3\tau) - f(t+2\tau), \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \Delta f(t+k\tau) &= f(t+(k+1)\tau) - f(t+k\tau). \end{aligned} \tag{34}$$

Может быть найдена первая разность для разности $\Delta f(t)$:

$$\Delta(\Delta f(t)) = \Delta f(t+\tau) - \Delta f(t) = (f(t+2\tau) - f(t+\tau)) - (f(t+\tau) - f(t)) = f(t+2\tau) - 2f(t+\tau) + f(t). \tag{35}$$

Эту разность обычно называют разностью второго порядка (или второй разностью) функции $f(t)$ и обозначают $\Delta^2 f(t)$.

Подобно этому определяются третья разность $\Delta^3 f(t)$ функции $f(t)$ и разности более высоких порядков. Разность n-го порядка функции $f(t)$ запишется в виде

$$\Delta^n f(t) = f(t+n\tau) - nf(t+(n-1)\tau) + \frac{n(n-1)}{2!} f(t+(n-2)\tau) + \cdots + (-1)^k C_n^k f(t+(n-k)\tau) + \cdots + (-1)^n f(t). \tag{36}$$

б) *Определение линейного разностного уравнения и начальных условий.*

Определение. *Линейным разностным уравнением порядка n для функции $f(t)$ называется уравнение вида*

$$a_0\Delta^n y(t) + a_1\Delta^{n-1} y(t) + \cdots + a_{n-1}\Delta y(t) + a_n y(t) = g_1(t). \tag{37}$$

Если $g_1(t) \equiv 0$, то уравнение (37) однородное, в противном случае — неоднородное. Учитывая записанное выше представление разностей, уравнению (37) можно придать иной вид:

$$b_0 y(t) + b_1 y(t+\tau) + b_2 y(t+2\tau) + \cdots + b_n y(t+n\tau) = g_1(t). \tag{38}$$

Полагая теперь $t = \omega\tau$, можно добиться того, что приращение аргумента станет равным 1. Поэтому в дальнейшем будем предпола-

гать, что $\tau = 1$. В этих условиях (считая $b_n \neq 0$), производя деление обеих частей уравнения (38) на b_n и обозначая

$$\frac{b_k}{b_n} = C_k \qquad (k = 0,\ 1,\ 2,\ \ldots,\ n), \qquad \frac{g_1(t)}{b_n} = g(t),$$

разностное уравнение порядка n теперь можно записать так:

$$y(t+n) + C_{n-1}y(t+(n-1)) + \cdots + C_1 y(t+1) + C_0 y(t) = g(t). \tag{39}$$

В случае разностного уравнения n-го порядка начальные условия представляют собой задание функции в n начальных интервалах (каждый длиной $\tau = 1$, т. е. задание функции на интервале $0 \leqslant t < n$).

Совместно с уравнением (39) эти условия дают возможность найти выражение $y(t+n)$, затем $y(t+(n+1))$ и т. д.

Однако такой путь решения довольно громоздкий. Операционное исчисление позволяет значительно упростить этот процесс.

в) *Решение разностного уравнения.* Покажем решение линейных разностных уравнений на примере разностного уравнения второго порядка.

Пусть дано уравнение

$$y(t+2) + C_1 y(t+1) + C_0 y(t) = g(t), \tag{40}$$

причем значения функции $y(t)$ заданы в интервале $0 \leqslant t < 2$.

Найдем для этого уравнения соответствующее уравнение в изображениях, полагая $y(t) \doteqdot Y(p)$, $g(t) \doteqdot G(p)$. По теореме упреждения имеем

$$y(t+1) \doteqdot e^p \left(Y(p) - \int_0^1 y(t)\, e^{-pt} dt \right),$$

$$y(t+2) \doteqdot e^{2p} \left(Y(p) - \int_0^2 y(t)\, e^{-pt} dt \right).$$

Поэтому уравнением в изображениях будет равенство

$$e^{2p}\left(Y(p) - \int_0^2 y(t)\, e^{-pt} dt \right) + C_1 e^p \left(Y(p) - \int_0^1 y(t)\, e^{-pt} dt \right) + C_0 Y(p) = G(p),$$

или

$$Y(p)(e^{2p} + C_1 e^p + C_0) = G(p) + C_1 e^p \int_0^1 y(t)\, e^{-pt} dt + e^{2p} \int_0^2 y(t)\, e^{-pt} dt.$$

Обозначим

$$C_1 e^p \int_0^1 y(t)\, e^{-pt} dt + e^{2p} \int_0^2 y(t)\, e^{-pt} dt = \Psi(p),$$

тогда

$$Y(p)(e^{2p}+C_1e^p+C_0)=G(p)+\Psi(p),$$

откуда

$$Y(p)=\frac{G(p)}{e^{2p}+C_1e^p+C_0}+\frac{\Psi(p)}{e^{2p}+C_1e^p+C_0}. \qquad (41)$$

Переходя к оригиналам, находим искомое решение.

2. Разностные уравнения с дискретным аргументом. В случае дискретного изменения аргумента значения t можно считать целочисленными. Тогда заданные и искомые функции смогут принимать лишь значения, соответствующие точкам $t=0$, $t=1$, $t=2$,... ..., т. е. будут образовывать последовательности. Индекс, указывающий номер члена последовательности, как это принято в теории разностных уравнений, обозначим через ν. Тогда для каждого значения ν можно записать соответствующее разностное уравнение

$$C_0y_\nu+C_1y_{\nu+1}+\cdots+C_{n-1}y_{\nu+n-1}+y_{\nu+n}=g_\nu(t) \qquad (42)$$

(не нарушая общности, считаем $C_n=1$, в противном случае выполняем деление обеих частей уравнения на этот коэффициент).

Уравнения вида (42) называют также рекуррентными уравнениями, так как, имея начальные значения y_0, y_1, y_2, ..., y_{n-1} и меняя последовательно значения ν, $\nu\in Z_0$, можно получить значения y_n, y_{n+1}, ...

Для решения уравнений вида (42) используется тот же прием, что и для решения уравнений с непрерывным аргументом. При этом, естественно, несколько изменяются расчетные формулы.

Примеры применения операционного исчисления в электротехнике. Наибольшее применение в электротехнике операционное исчисление получило при исследовании переходных процессов в линейных цепях с сосредоточенными параметрами r, L и C. Это объясняется тем, что явления, происходящие в таких цепях, описываются обыкновенными линейными дифференциальными уравнениями и их системами, которые легко решаются с помощью операционного исчисления.

Прежде чем рассматривать конкретные примеры, ознакомимся с основными особенностями и законами переходных процессов в электрических цепях.

Переходным процессом называется явление, наблюдающееся в цепи при переходе от одного установившегося режима к другому.

Переходные процессы возникают в электрических цепях в результате коммутаций (включения или выключения э. д. с., различных переключений, короткого замыкания в цепи, внезапного изменения параметров в цепи и т. д.). Эти процессы в электрических цепях всегда являются электромагнитными. Они протекают обычно с очень большой скоростью и, как правило, заканчиваются по истечении долей секунды. При этом возможны случаи, когда

напряжения и токи в цепи или на отдельных ее элементах при переходном процессе значительно превосходят их значения в установившемся режиме. Последнее может привести к выходу из строя некоторых элементов цепи. При протекании переходных процессов в электрических цепях всегда выполняются законы коммутации (законы переходных процессов). Они гласят:

а) *Ток в индуктивности L не может измениться скачком.* В начальный момент (непосредственно после коммутации) он сохраняет то значение, которое было в момент, непосредственно предшествующий коммутации: $i_L(0_+) = i_L(0_-) = i_L(0)$.

б) *Напряжение на емкости C не может измениться скачком.* В начальный момент (непосредственно после коммутации) оно сохраняет то значение, которое было в момент, непосредственно предшествующий коммутации: $u_C(0_+) = u_C(0_-) = u_C(0)$.

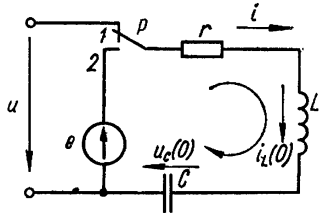

Рис. 56

Значения токов в индуктивностях и напряжений на обкладках конденсаторов в момент времени, непосредственно предшествующий коммутации в цепи, $i_L(0)$ и $u_C(0)$, определяют начальные условия переходного процесса. При расчете переходного процесса в электрической цепи эти условия необходимо выявить до выполнения всех остальных вычислений. Если все $i_L(0)$ и $u_C(0)$ равны нулю, то в цепи имеют место нулевые начальные условия, а токи в индуктивностях и напряжения на конденсаторах в переходном процессе начнут изменяться от нулевых значений. При ненулевых начальных условиях для определения знаков $i_L(0)$ и $u_C(0)$ надо задаться направлениями обхода контуров цепи, в которых будет происходить переходный процесс. Положительные знаки $i_L(0)$ и $u_C(0)$ сохранятся, если их направления совпадают с направлением обхода контура. В противном случае знаки $i_L(0)$ и $u_C(0)$ изменятся на противоположные. Здесь токи в индуктивностях и напряжения на конденсаторах в переходном процессе начнут изменяться от тех значений, которые они имели в момент, непосредственно предшествующий коммутации (с учетом установленных знаков соответствующих величин).

Пусть в электрической цепи, изображенной на рис. 56, рубильник P переключается из положения *1* в положение *2*. Тогда в контуре r, L, C возникнет переходный процесс. Примем, что его начальные условия ненулевые: $i_L(0) \neq 0$ и $u_C(0) \neq 0$. При направлениях тока в индуктивности и напряжения на обкладках конденсатора в начальный момент переходного процесса, показанных на рис. 56, и выбранном направлении обхода контура имеем $i_L(0) > 0$ и $u_C(0) > 0$. Возьмем направление мгновенного значения тока переходного процесса $i = i(t)$, совпадающим с направлением обхода контура. Так как направление источника э. д. с. $e = e(t)$, действующего в контуре r, L, C во время переходного процесса,

совпадает с направлением обхода этого контура, то по второму закону Кирхгофа получаем уравнение

$$ri + L\frac{di}{dt} + \frac{1}{C}\int_0^t idt + u_C(0) = e. \qquad (43)$$

Обозначим: $i(t) = i \doteqdot I(p)$ — изображение тока переходного процесса в контуре; $e(t) = e \doteqdot E(p)$ — изображение внешней э. д. с., действующей в контуре. Тогда уравнение цепи r, L, C (43) в операторной форме примет вид

$$rI(p) + L(pI(p) - i_L(0)) + \frac{1}{pC}I(p) + \frac{u_C(0)}{p} = E(p). \qquad (44)$$

Это уравнение можно записать так:

$$\left(r + pL + \frac{1}{pC}\right)I(p) = E(p) + Li_L(0) - \frac{u_C(0)}{p},$$

откуда находим выражение для изображения тока переходного процесса в виде

$$I(p) = \frac{E(p) + Li_L(0) + \dfrac{-u_C(0)}{p}}{r + pL + \dfrac{1}{pC}}. \qquad (45)$$

Полученная зависимость представляет собой закон Ома в операторной форме. Его можно записать так:

$$I(p) = \frac{\mathcal{E}(p)}{Z(p)}, \qquad (46)$$

где $\mathcal{E}(p) = E(p) + Li_L(0) + \dfrac{-u_C(0)}{p}$ — изображение всех (внешних и внутренних) э. д. с., действующих в контуре; $Z(p) = r + pL + \dfrac{1}{pC}$ — операторное сопротивление контура r, L, C; $\dfrac{-u_C(0)}{p}$ — изображение начальной э. д. с. емкости (включая знак «минус»), уравновешивающей начальное напряжение на обкладках конденсатора и направленной навстречу $u_C(0)$.

Операторное сопротивление контура r, L, C имеет вид

$$Z(p) = r + pL + \frac{1}{pC}. \qquad (47)$$

Оно может быть получено из выражения комплекса полного сопротивления этого контура

$$Z = r + j\omega L + \frac{1}{j\omega C}$$

путем замены $j\omega$ на p (в электротехнике символ i ($i^2 = -1$) обозначают буквой j).

Закон Ома в операторной форме позволяет непосредственно исследовать переходные процессы только в неразветвленных электри-

ческих цепях. При рассмотрении переходных процессов в разветвленных и сложных электрических цепях необходимо использовать первый и второй законы Кирхгофа, которые имеют в операторной форме следующий вид:

$$\text{первый закон} — \sum_{k=1}^{n} I_k(p) = 0, \tag{48}$$

$$\text{второй закон} — \sum_{k=1}^{m} Z_k(p) I_k(p) \quad \sum_{k=1}^{l} \mathcal{E}_k(p). \tag{49}$$

При составлении уравнений цепи по этим законам «правила знаков» остаются такими же, как и при расчете установившихся режимов в электрических цепях постоянного и переменного тока. В частности, если мгновенное значение тока переходного процесса $i_n(t)$, протекающего в ветви n, принято направленным к заданному узлу (для которого составляется уравнение по первому закону Кирхгофа), то изображение этого тока $I_n(p)$ берется с одним знаком (например, со знаком «плюс»). Если же ток $i_m(t)$ направлен от узла, то его изображение $I_m(p)$ берется с другим знаком (со знаком «минус»).

Составляя уравнения по второму закону Кирхгофа, здесь надо помнить, что кроме внешних э. д. с. $e_k(t) = e_k$ в контурах, содержащих индуктивности и емкости при ненулевых начальных условиях, будут действовать еще и внутренние э. д. с. (начальные э. д. с. самоиндукции и емкости). Причем, если направление $i_{L_k}(0)$ совпадает с направлением обхода контура, то слагаемое $L_k i_{L_k}(0)$ следует брать со знаком «плюс», если же и $u_{C_k}(0)$ направлено по обходу контура, то результирующий знак слагаемого $\frac{u_C(0)}{p}$ должен быть «минус», так как начальная э. д. с. емкости всегда направлена навстречу начальному напряжению на обкладках конденсатора $u_C(0)$.

Рассматривая зависимости (46), (48) и (49), замечаем, что законы Ома и Кирхгофа в операторной форме имеют тот же вид, что и при установившихся режимах в цепях постоянного и переменного тока. Поэтому, применяя операционное исчисление для расчета переходных процессов, в принципе можно использовать все методы расчета сложных линейных электрических цепей с постоянными параметрами. При исследовании переходных процессов в сложных и разветвленных электрических цепях (в последнем случае при ненулевых начальных условиях) наибольшее применение получили *метод уравнений Кирхгофа, метод контурных токов и метод наложения.* При расчете переходных процессов в неразветвленных цепях, а также в простых разветвленных цепях при нулевых начальных условиях применяется *закон Ома в операторной форме.* При этом в разветвленной цепи непосредственно определяется только ток переходного режима в ветви, содержащей источник э. д. с. (вся цепь первоначально сводится к простой неразветвленной цепи).

Во всех случаях расчета переходных процессов в электрических цепях операторным методом сохраняется такая последовательность операций: сначала определяются начальные условия, затем записывается уравнение или система уравнений для заданной цепи в операторной форме, что позволяет найти изображения искомых токов или напряжений. И, наконец, по полученным изображениям отыскиваются оригиналы — мгновенные значения токов или напряжений переходного режима.

Применение операционного исчисления к расчету переходных процессов в линейных электрических цепях с сосредоточенными постоянными параметрами покажем на примерах.

Примеры

1. Найти переходные значения тока и напряжений (i, u_r, u_C) в цепи, изображенной на рис. 57, при переключении рубильника P из положения *1* в положение *2*, если $U = 100$ В, $r = 100$ Ом, $C = 10$ мкФ.

Р е ш е н и е. Будем считать, что до переключения рубильника из положения *1* в положение *2* конденсатор C был заряжен до напряжения источника U. Если обход контура взять совпадающим с ходом стрелки часов, то начальное напряжение на конденсаторе u_C (0) следует считать положительным: $u_C(0) = U = 100$ В.

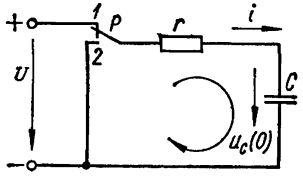

Рис. 57

Операторное сопротивление контура

$$Z(p) = r + \frac{1}{pC}.$$

Внешних источников э. д. с. в контуре нет. Изображение внутренней э. д. с. (начальной э. д. с. емкости) будет

$$\frac{-u_C(0)}{p} = \frac{-U}{p} = \mathcal{E}(p).$$

По закону Ома в операторной форме имеем

$$I(p) = \frac{\mathcal{E}(p)}{Z(p)} = \frac{\dfrac{-U}{p}}{r + \dfrac{1}{pC}} = \frac{-UC}{rCp + 1}.$$

Изображение тока переходного процесса удовлетворяет условиям применения второй теоремы разложения. Поэтому можно записать

$$I(p) = \frac{-UC}{rCp + 1} = \frac{P(p)}{Q(p)},$$

где

$$P(p) = -UC = -100 \cdot 10 \cdot 10^{-6} = -10^{-3};$$

$$Q(p) = rCp + 1; \quad \alpha_1 = -\frac{1}{rC} = -\frac{1}{100 \cdot 10 \cdot 10^{-6}} = -10^3;$$

$$P(\alpha_1) = -10^{-3};$$

$$Q'(p) = rC = 100 \cdot 10 \cdot 10^{-6} = 10^{-3};$$

$$Q'(\alpha_1) = 10^{-3}.$$

В результате получаем мгновенное значение тока переходного процесса в виде

$$i = i(t) = \frac{P(\alpha_1)}{Q'(\alpha_1)} e^{\alpha_1 t} = \frac{-10^{-3}}{10^{-3}} e^{-10^3 t} = -e^{-1000t} \text{ А}.$$

Отрицательный знак тока означает, что при разряде конденсатора C через активное сопротивление r ток $i\,(t)$ направлен в сторону, противоположную направлению обхода контура, выбранного нами раньше. Мгновенное значение напряжения на активном сопротивлении в переходном процессе будет также отрицательным:

$$u_r = ri = -100e^{-1000t} \text{ В},$$

а мгновенное значение переходного напряжения на конденсаторе, определяемое по второму закону Кирхгофа

$$ri + u_C = 0, \quad u_C = -ri = -u_r$$

получается положительным: $u_C = 100e^{-1000t}$ В.

Характер изменения i и u_C в переходном режиме показан на рис. 58.

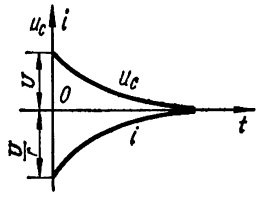

Рис. 58

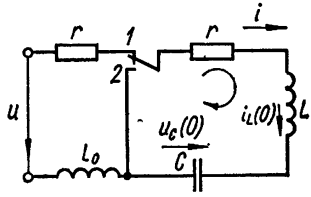

Рис. 59

2. Найти переходные значения тока и напряжений (i, u_r, u_C, u_L) в цепи, изображенной на рис. 59, если переключение рубильника из положения *1* в положение *2* происходит в момент времени $t = 0$ и $u = 141 \sin(2500t + 105°)$ В, $r = 1000$ Ом, $L = 1$ Г, $L_0 = 0{,}2$ Г, $C = 0{,}4$ мкФ.

Р е ш е н и е. До переключения рубильника в цепи протекал синусоидальный ток

$$i = I_m \sin(2500t + 105° - \varphi) \text{ А},$$

а на обкладках конденсатора было напряжение

$$u_C = x_C I_m \sin(2500t + 105° - \varphi - 90°) \text{ В}.$$

Здесь $\omega = 2500$ 1/с — угловая частота напряжения источника электрической энергии;

$\varphi = \operatorname{arctg} \dfrac{x}{2r}$ — угол сдвига фаз между напряжением источника и током в цепи;

$$x = x_L - x_C = \omega(L + L_0) - \frac{1}{\omega C} = 2500(1 + 0{,}2) - \frac{10^6}{2500 \cdot 0{,}4} = 3000 - 1000 = 2000 \text{ Ом}$$

— цепь имеет индуктивный характер, причем $x_C = 1000$ Ом;

$$\varphi = \operatorname{arctg} \frac{x}{2r} = \operatorname{arctg} \frac{2000}{2 \cdot 1000} = \operatorname{arctg} 1 = 45°;$$

$$I_m = \frac{U_m}{z}; \quad U_m = 141 \text{ В};$$

$$z = \sqrt{(2r)^2 + x^2} = \sqrt{(2000)^2 + (2000)^2} = 2\sqrt{2} \cdot 10^3 \text{ Ом};$$

$$I_m = \frac{141}{2\sqrt{2} \cdot 10^3} = 0{,}05 \text{ А};$$

$$x_C I_m = 1000 \cdot 0{,}05 = 50 \text{ В}.$$

Поэтому имеем

$$i = 0{,}05 \sin (2500t + 105° - 45°) = 0{,}05 \sin (2500t + 60°) \text{ А},$$

$$u_G = 50 \sin (2500t + 105° - 45° - 90°) = 50 \sin (2500t - 30°) \text{ В}.$$

При $t = 0$ (в момент, непосредственно предшествующий коммутации) получаем

$$i_L (0) = 0{,}05 \sin 60° = 0{,}0433 \text{ А},$$

$$u_G (0) = 50 \sin (-30°) = -25 \text{ В}.$$

На рис. 59 показано истинное направление напряжения $u_G (0)$.

После переключения рубильника переходный процесс будет протекать в контуре, содержащем сопротивление r, индуктивность L и емкость C. Внешнего источника э. д. с. в контуре нет. Выберем обход контура по ходу стрелки часов. Тогда имеем $i_L (0) = 0{,}0433$ А и $u_G = -25$ В, а изображения внутренних э. д. с. запишутся в виде (обратите внимание на знаки):

$$\mathcal{E} (p) = Li_L (0) + \frac{-u_C (0)}{p} = 0{,}0433 + \frac{25}{p}.$$

Операторное сопротивление контура $Z (p) = r + pL + \dfrac{1}{pC}$. Закон Ома в операторной форме примет вид

$$I (p) = \frac{0{,}0433 + \dfrac{25}{p}}{r + pL + \dfrac{1}{pC}} = \frac{C (0{,}0433p + 25)}{LCp^2 + rCp + 1}.$$

Применим вторую теорему разложения, так как нами получена правильная рациональная и несократимая дробь с простыми корнями знаменателя.

Здесь

$$P (p) = C (0{,}0433p + 25) = 0{,}4 \cdot 10^{-6} (0{,}0433p + 25);$$

$$Q (p) = LCp^2 + rCp + 1 = 1 \cdot 0{,}4 \cdot 10^{-6} p^2 + 1000 \cdot 0{,}4 \cdot 10^{-6} p + 1 =$$

$$= 0{,}4 \cdot 10^{-6} p^2 + 0{,}4 \cdot 10^{-3} p + 1;$$

$$\alpha_1 = -500 + j1500, \quad \alpha_2 = -500 - j1500;$$

$$Q' (p) = 0{,}4 \cdot 10^{-6} (2p + 1000);$$

$$P (\alpha_1) = 0{,}4 \cdot 10^{-6} (3{,}35 + j64{,}95);$$

$$P (\alpha_2) = 0{,}4 \cdot 10^{-6} (3{,}35 - j64{,}95);$$

$$Q' (\alpha_1) = 0{,}4 \cdot 10^{-6} j3000, \quad Q' (\alpha_2) = -0{,}4 \cdot 10^{-6} j3000.$$

При этом

$$i = i (t) = \frac{0{,}4 \cdot 10^{-6} (3{,}35 + j64{,}95)}{0{,}4 \cdot 10^{-6} j3000} e^{(-500 + j1500)t} +$$

$$+ \frac{0{,}4 \cdot 10^{-6} (3{,}35 - j64{,}95)}{-0{,}4 \cdot 10^{-6} j3000} e^{(-500 - j1500)t} =$$

$$= e^{-500t} \left(\frac{3{,}35 + j64{,}95}{j3000} e^{j1500t} - \frac{3{,}35 - j64{,}95}{j3000} e^{-j1500t} \right) =$$

$$= e^{-500t} \left(\frac{3{,}35}{1500} \left(\frac{e^{j1500t} - e^{-j1500t}}{2j} \right) + \frac{64{,}95}{1500} \left(\frac{e^{j1500t} + e^{-j1500t}}{2} \right) \right) =$$

$$= \frac{1}{1500} e^{-500t} (3{,}35 \sin 1500t + 64{,}95 \cos 1500t) = \frac{\sqrt{3{,}35^2 + 64{,}95^2}}{1500} e^{-500t} \times$$

$$\times \left(\frac{3{,}35}{\sqrt{4230}} \sin 1500t + \frac{64{,}95}{\sqrt{4230}} \cos 1500t \right) = 0{,}0434 e^{-500t} \sin (1500t + 87° 5') \text{ А}.$$

Таким образом, мгновенное значение тока переходного процесса равно:

$$i = 0{,}0434 e^{-500t} \sin (1500t + 87° 5') \text{ А}.$$

Найдем переходные напряжения на активном сопротивлении и на индуктивности контура:

$$u_r = ri = 1000 \cdot 0{,}0434 e^{-500t} \sin (1500t + 87° 5') =$$

$$= 43{,}4 e^{-500t} \sin (1500t + 87° 5') \text{ В};$$

$$u_L = L \frac{di}{dt} = L \cdot 0{,}0434 \cdot 500 e^{-500t} (3 \cos (1500t + 87° 5') - \sin (1500t + 87° 5')) =$$

$$= 1 \cdot 21{,}7 e^{-500t} \sqrt{10} \left(\frac{3}{\sqrt{10}} \cos (1500t + 87° 5') - \frac{1}{\sqrt{10}} \sin (1500t + 87° 5') \right) =$$

$$= 68{,}7 e^{-500t} \sin (1500t + 87° 5' + 108° 27') = 68{,}7 e^{-500t} \sin (1500t + 195° 32') \text{ В}.$$

Теперь можно определить по второму закону Кирхгофа и мгновенное значение напряжения на обкладках конденсатора в переходном режиме:

$$u_C = - u_r - u_L = - 43{,}4 e^{-500t} \sin (1500t + 87° 5') -$$

$$- 68{,}7 e^{-500t} \sin (1500t + 195° 32') = 68{,}7 e^{-500t} \sin (1500t - 21° 22') \text{ В}.$$

Выполним проверку расчета для момента времени $t = 0$. Получим

$$i(0) = i_L(0) = 0{,}0434 \sin 87° 5' = 0{,}0433 \text{ А};$$

$$u_C(0) = 68{,}7 \sin (- 21° 22') = - 25 \text{ В},$$

что подтверждает правильность полученных результатов.

На рис. 60 показан характер изменения переходных тока в контуре и напряжения на индуктивности. Как видим, $i(t)$ и $u_L(t)$ представляют собой периодические синусоидальные колебания с убывающей амплитудой.

Аналогичный характер изменения будет и у переходного напряжения на емкости $u_C(t)$: частота этих колебаний равна 1500 1/с, а их амплитуда уменьшается по показательному закону с коэффициентом затухания 500 1/с.

3. Найти переходное значение тока в конденсаторе при переключении рубильника из положения *1* в положение *2*, когда в цепи, изображенной на рис. 61, известны величины U, r_1, r_2, C и U_0.

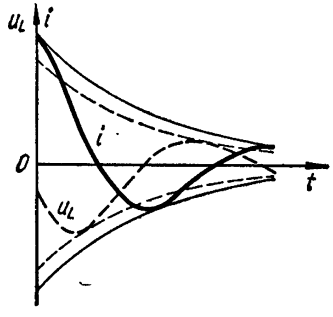

Рис. 60

Р е ш е н и е. Здесь при переходном режиме мы имеем разветвленную цепь при ненулевых начальных условиях. Поэтому

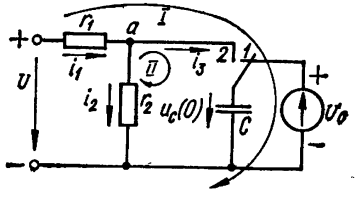

Рис. 61

воспользоваться для расчета законом Ома в операторной форме нельзя, а надо применить один из методов расчета сложных электрических цепей. Пусть это будет метод уравнений Кирхгофа. Будем считать, что конденсатор до переходного процесса был заряжен до напряжения U_0.

Тогда для узла a, I и II контуров в операторной форме получим систему уравнений

$$\begin{cases} I_1(p) = I_2(p) + I_3(p), \\ r_1 I_1(p) + \dfrac{1}{pC} I_3(p) = \dfrac{U}{p} - \dfrac{U_0}{p}, \\ -r_2 I_2(p) + \dfrac{1}{pC} I_3(p) = -\dfrac{U_0}{p}. \end{cases}$$

Решая эту систему относительно $I_3(p)$, найдем

$$I_3(p) = \frac{C(r_2 U - (r_1 + r_2) U_0)}{r_1 r_2 C p + (r_1 + r_2)},$$

откуда, после применения теоремы разложения, имеем

$$i_G = i_3 = \left(\frac{U}{r_1} - \left(\frac{1}{r_1} + \frac{1}{r_2}\right) U_0\right) e^{-\frac{r_1 + r_2}{r_1 r_2 C} t}.$$

Аналогично могут быть найдены токи i_1 и i_2.

Из приведенных примеров видно, что при расчете переходных процессов в электрических цепях операторным методом отпадает необходимость специального определения постоянных интегрирования. Это существенно упрощает решение задачи.

Дискретное преобразование Лапласа и z-преобразование

Названные преобразования применяются в теории автоматического управления и в радиотехнике при рассмотрении импульсных систем. Ознакомимся с этими преобразованиями.

Пусть $f(t)$ — комплексная функция действительного аргумента t, определенная для всех $t \geqslant 0$.

Последовательность $\{f(n)\}$ обозначается кратко $f(n)$ и называется решетчатой функцией ($f(t)$ для решетчатой функции $f(n)$ называется порождающей функцией).

Аргумент решетчатой функции принимает только целые значения ($n \in Z_0$). При отрицательных значениях аргумента решетчатая функция равна нулю. Интегрировать решетчатую функцию нельзя, поэтому вместо интеграла Лапласа для нее составляется бесконечный ряд $\sum_{n=0}^{\infty} f(n) e^{-np}$, где p — комплексная переменная. Если этот ряд сходится, то выписанную выше операцию суммирования называют дискретным преобразованием Лапласа решетчатой функции $f(n)$. Функцию $f(n)$ называют оригиналом, а сумму $F^*(p)$ ряда ($F^*(p) = \sum_{n=0}^{\infty} f(n) e^{-np}$) — изоб-

ражением. Это записывают так: $f(n) \doteqdot F^*(p)$ или $F^*(p) \doteqdot f(n)$ (иногда вместо символа соответствия $\doteqdot$ для дискретного преобразования Лапласа употребляют символ $\div$).

Всякая функция $f(t)$, для которой существует обычное преобразование Лапласа, порождает решетчатую функцию, для которой, в свою очередь, определено дискретное преобразование Лапласа.

Свойства дискретного преобразования Лапласа решетчатой функции, в основном, такие же, как и обычного преобразования Лапласа, однако здесь во всех случаях интегралы заменяются бесконечными суммами.

В теории автоматического управления рассматриваются дискретные функции, заданные равноотстоящими ординатами. Такую функцию $x[nT_n]$ можно представить в виде суммы последовательностей δ-функций:

$$x[nT_n] = \sum_{n=0}^{\infty} x_n \delta(t - nT_n),$$

где $x[nT_n] = \{0, T_1, 2T_2, 3T_3, \ldots, nT_n, \ldots\}$ — решетчатая функция.

Записанное представление функции $x[nT_n]$ основано на том, что $\int_a^b f(t)\,\delta(t - t_0)\,dt = f(t_0)$, причем интегрирование заменено суммированием.

Применим к функции $x[nT_n]$ преобразование Лапласа, учитывая, что $\delta(t) \doteqdot 1$, и потому

$$\delta(t - nT_n) \doteqdot e^{-pnT_n} = (e^{pT_n})^{-n}.$$

Получим

$$x[nT_n] \doteqdot \sum_{n=0}^{\infty} x_n (e^{pT_n})^{-n}.$$

Функцию e^{pT_n} комплексной переменной p обозначим через z, т. е. $z = e^{pT_n}$. Тогда

$$x[nT_n] \doteqdot \sum_{n=0}^{\infty} x_n z^{-n}.$$

Сумма полученного здесь ряда — это функция комплексной переменной z. Ее обозначают $X^*(z)$.

Операцию суммирования для получения $X^*(z)$ в теории автоматического управления называют z-преобразованием функции времени $[nT_n]$ в функцию $X^*(z)$ комплексной переменной z.

Функция $x[nT_n]$ — оригинал, а функция $X^*(z)$ — изображение.

Понятия z-преобразования и дискретного преобразования Лапласа для прямоугольного импульса совпадают. В других случаях такого совпадения нет. Отыскание оригиналов по изображениям для дискретного преобразования Лапласа и z-преобразования

Таблица 3

$f(t)$	$F(p)$	$f[nT_n]$	$F^*(z)$
$\delta(t)$	1	$\delta[nT_n]$	1
$1(t)$	$\frac{1}{p}$	$1[nT_n]$	$\frac{z}{z-1}$
t	$\frac{1}{p^2}$	nT_n	$T_n\frac{z}{(z-1)^2}$
$\frac{1}{2}t^2$	$\frac{1}{p^3}$	$\frac{1}{2}[nT_n]$	$\frac{T_n^2 z(z+1)}{2(z-1)^2}$
$1-e^{-\alpha t}$	$\frac{\alpha}{p(p+\alpha)}$	$1-e^{-\alpha nT_n}$	$\frac{(1-e^{-\alpha T_n})z}{(z-1)(z-e^{-\alpha T_n})}$
$\sin\omega t$	$\frac{\omega}{p^2+\omega^2}$	$\sin(\omega nT_n)$	$\frac{z\sin(\omega T_n)}{z^2-2z\cos(\omega T_n)+1}$
$\cos\omega t$	$\frac{p}{p^2+\omega^2}$	$\cos(\omega nT_n)$	$\frac{z(z-\cos(\omega T_n))}{z^2-2z\cos(\omega T_n)+1}$

производится по формулам, подобным формулам, применяемым в случае обычного преобразования Лапласа. При использовании этих формул интегрирование, как и обычно, выполняется в комплексной плоскости.

В таблице 3 приведены для сравнения некоторые формулы преобразований Лапласа для непрерывных функций и z-преобразований решетчатых функций.

§ 20. СВЯЗЬ ИНТЕГРАЛА ЛАПЛАСА С ИНТЕГРАЛОМ ФУРЬЕ. ФОРМУЛА ОБРАЩЕНИЯ

Если функция $f(t)$ удовлетворяет требованиям, предъявляемым к оригиналу (это значит, что она и абсолютно интегрируема при всех $t \in]-\infty; +\infty[$) и, кроме того, имеет лишь конечное число экстремумов в каждом промежутке, то для нее можно записать интеграл Фурье, причем имеют место формулы:

$$\frac{f(t+0)+f(t-0)}{2} = \frac{1}{2\pi}\int_{-\infty}^{\infty} C(\omega)\, e^{i\omega t} d\omega,$$

$$C(\omega) = \int_{-\infty}^{\infty} f(t)\, e^{t\omega t} dt.$$

Учитывая, что в интеграле Лапласа параметр $p = \sigma + i\omega$ ($\operatorname{Re} p = \sigma$) и для сходимости интеграла выбирается $\sigma > S_0$, а

$f(t)|_{t<0} \equiv 0$, можно записать

$$F(p) = F(\sigma + i\omega) = \int_0^{\infty} f(t)\, e^{-pt} dt = \int_{-\infty}^{\infty} f(t)\, e^{-\sigma t} e^{-i\omega t} dt.$$

Сравнивая записанный интеграл Лапласа с преобразованием Фурье, устанавливают, что изображение $F(\sigma + i\omega) = F(p)$ есть прямое преобразование Фурье для функции $\varphi(t) = f(t)\, e^{-\sigma t}$ ($F(p) = \int_{-\infty}^{\infty} \varphi(t)\, e^{-i\omega t}\, dt$). Этот факт используют для отыскания по известному изображению функции $\varphi(t)$, а затем и оригинала $f(t)$. Очевидно,

$$\frac{f(t+0) + f(t-0)}{2} e^{-\sigma t} = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\sigma + i\omega)\, e^{i\omega t} dt.$$

Заменяя переменную интегрирования ω на переменную p по формуле $p = \sigma + i\omega$ (σ фиксировано), получают

$$i\omega = p - \sigma, \quad \omega = \frac{1}{i}(p - \sigma), \quad d\omega = \frac{dp}{i}.$$

Пределы интегрирования записывают в виде $\sigma - i\infty$ и $\sigma + i\infty$ (или $\sigma - i\omega$ и $\sigma + i\omega$), учитывая, что переменная p проходит все комплексные значения с вещественной частью σ и всеми возможными мнимыми частями.

Получают формулу

$$\frac{f(t+0) + f(t-0)}{2} = \frac{1}{2\pi i} \int_{\sigma - i\infty}^{\sigma + i\infty} F(p)\, e^{pt} dp. \qquad (50)$$

Интегрирование ведется вдоль прямой $\sigma = \text{const}$, параллельной мнимой оси. Формула (50) определяет обратное преобразование Лапласа и называется формулой обращения Римана—Мелллина.

ГЛАВА V

ТЕОРИЯ ПОЛЯ

§ 21. ПОНЯТИЕ ПОЛЯ

Полем называется область пространства, каждой точке P которой поставлена в однозначное соответствие некоторая величина V (P) (иногда полем называют функцию V (P)).

Если величина $V(P)$ является физической, то поле называется физическим. При этом, в зависимости от природы функции

$V(P)$, поля разделяют на скалярные и векторные. Примерами скалярных физических полей могут быть поля температуры, атмосферного давления, плотности воздуха, электрического потенциала и др. К векторным физическим полям относятся, например, поля силы тяжести, скорости частиц текущей жидкости, сдвига точек упругого тела, магнитной индукции, плотности электрического тока.

Если функция $V(P)$ не изменяется с течением времени, то поле называется стационарным, или постоянным, в противном случае,— нестационарным, или переменным.

Для получения общих результатов, справедливых для любых конкретных физических полей, всякому физическому полю ставится в соответствие его математическая модель, где абстрагируются от заданной физической величины, заменяя ее математическим понятием функции, а точке пространства приписывают координаты, считая их аргументами.

Математическая теория поля изучает свойства векторных и скалярных полей, к рассмотрению которых приводят многие задачи физики, электротехники, математики и других наук. При этом функцию $V(P)$ часто рассматривают как функцию векторного аргумента r (радиус-вектора точки P).

Математическим аппаратом, наиболее соответствующим изучаемым в теории поля явлениям, служат векторная алгебра и векторный анализ. Обобщением теории скалярных и векторных полей явилась теория тензорных полей, которая играет важную роль, например, в общей теории относительности.

§ 22. СКАЛЯРНОЕ ПОЛЕ. ГРАДИЕНТ

Поверхности и линии уровня. Для задания скалярного поля надо взять скалярную функцию $U(P) = U(x, y, z)$. Введем понятие поверхности уровня скалярного поля.

Определение. *Поверхностью уровня скалярного поля $U(P)$ называется такая поверхность, на которой функция $U(P)$ имеет постоянное значение.*

Уравнение поверхности уровня: $U(x, y, z) = C$, где C — некоторая постоянная.

При различных значениях постоянной C получаем семейство поверхностей уровня. Будем считать, что функция $U(x, y, z)$, задающая поле, однозначна. Тогда, очевидно, через каждую точку поля проходит лишь одна поверхность уровня. Примерами поверхностей уровня являются поверхности равных температур в некотором теле и поверхности равного потенциала V в электрическом поле $V(x, y, z) = C$.

Для плоского скалярного поля рассматривают линии уровня; на них функция, задающая поле, имеет постоянное значение. Такими линиями являются, например, на топографических картах горизонтали — линии, вдоль которых высота местности над услов-

ным уровнем моря постоянна. Линии уровня (линии равного потенциала или эквипотенциальные линии) находят широкое применение при рассмотрении плоскопараллельных электрических полей. Здесь они являются линиями, вдоль которых электрический потенциал во всех их точках постоянен. Это физически означает, что работа, совершаемая силами электрического поля по переносу единичного положительного заряда из любой точки данной линии уровня в точку, потенциал которой принят равным нулю (в точку нулевого потенциала), будет одинакова. Совокупность линий равного потенциала дает наглядное изображение электрического поля, что облегчает его изучение.

Линии уровня применяются в математике при исследовании поверхностей (методом сечений). Уравнение линий уровня: $U(x, y) = C$.

Примеры

1. Найти поверхность уровня поля $U = x^2 + y^2 + z^2$, проходящую через точку P_0 (1, —1, 2).

Р е ш е н и е. Уравнение поверхности уровня

$$U = C; \quad x^2 + y^2 + z^2 = C.$$

Очевидно, в данном случае $C \geqslant 0$. Поверхности уровня — семейство сфер с центром в начале координат. Из этого семейства надо выбрать сферу, проходящую через точку P_0 (1, —1, 2).

Подставим координаты этой точки в уравнение поверхностей уровня:

$$1^2 + (-1)^2 + 2^2 = C, \quad C = 6.$$

Уравнение искомой поверхности уровня имеет вид

$$x^2 + y^2 + z^2 = 6.$$

Это сфера радиуса $R = \sqrt{6}$ с центром в начале координат.

2. Найти линии уровня поля $U = x^2 - y^2$.

Р е ш е н и е. Положим: $U = C$; $x^2 - y^2 = C$.

При $C > 0$ линии уровня — гиперболы с вершинами на оси Ox, при $C < 0$ — гиперболы с вершинами на оси Oy. Если $C = 0$, получим прямые $y = \pm x$ — асимптоты этих гипербол.

Производная по направлению. Пусть дано скалярное поле $U = U(P)$. Возьмем в пространстве, где задано это поле, некоторую точку P и проведем через нее какое-нибудь направление $\boldsymbol{l}$ (рис. 62). Пусть точка P_1 также находится на направлении $\boldsymbol{l}$. Найдем приращение функции U при переходе из точки P в точку P_1:

$$\Delta U = U(P_1) - U(P).$$

Рассмотрим

$$\lim_{P_1 \to P} \frac{\Delta U}{PP_1} = \lim_{P_1 \to P} \frac{U(P_1) - U(P)}{PP_1}$$

(предполагается, что точка P_1 неограниченно приближается к точке P, оставаясь на направлении $\boldsymbol{l}$).

Если этот предел существует, то он называется п р о и з в о д н о й от функции U по направлению $\boldsymbol{l}$ и обозначается $\frac{\partial U}{\partial l}$. Таким

образом,

$$\frac{\partial U}{\partial l} = \lim_{P_1 \to P} \frac{\Delta U}{PP_1}.$$

Если $\frac{\partial U}{\partial l} > 0$, то функция U в направлении l возрастает, если $\frac{\partial U}{\partial l} < 0$, то функция U в направлении l убывает.

Величина $\left|\frac{\partial U}{\partial l}\right|$ характеризует скорость изменения функции U в направлении l: чем больше $\left|\frac{\partial U}{\partial l}\right|$, тем быстрее изменяется (растет или убывает, в зависимости от знака $\frac{\partial U}{\partial l}$) функция U в направлении l. В этом состоит физический смысл производной по направлению.

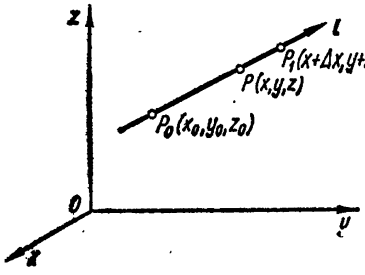

Рис. 62

Понятие производной по направлению является обобщением понятия частных производных $\frac{\partial U}{\partial x}$, $\frac{\partial U}{\partial y}$, $\frac{\partial U}{\partial z}$. Эти частные производные можно рассматривать, как производные от функции U по направлению соответствующих осей координат.

Выведем формулу для нахождения производной по направлению.

Пусть $\cos\alpha$, $\cos\beta$, $\cos\gamma$ — направляющие косинусы направления l, $P_0(x_0, y_0, z_0)$ — некоторая фиксированная точка на этом направлении. Очевидно, координаты точки $P(x, y, z)$ на направлении l можно записать в виде

$$\left.\begin{aligned} x &= x_0 + t\cos\alpha, \\ y &= y_0 + t\cos\beta, \\ z &= z_0 + t\cos\gamma, \end{aligned}\right\} \qquad (1)$$

где t — длина отрезка P_0P.

В точках $P(x, y, z)$, лежащих на l, функция $U(x, y, z) = U(x(t), y(t), z(t)) = \varphi(t)$ является функцией одной переменной t. Поэтому

$$\frac{\partial U}{\partial l} = \frac{d\varphi(t)}{dt}.$$

Применяя формулу дифференцирования сложной функции, находим

$$\frac{d\varphi(t)}{dt} = \frac{\partial U}{\partial x}\cdot\frac{dx}{dt} + \frac{\partial U}{\partial y}\cdot\frac{dy}{dt} + \frac{\partial U}{\partial z}\cdot\frac{dz}{dt}.$$

Но из формул (1) следует, что

$$\frac{dx}{dt} = \cos\alpha, \quad \frac{dy}{dt} = \cos\beta, \quad \frac{dz}{dt} = \cos\gamma$$

(x_0, y_0, z_0, $\cos\alpha$, $\cos\beta$, $\cos\gamma$ не зависят от переменной t).

Следовательно,

$$\frac{\partial U}{\partial l} = \frac{\partial U}{\partial x}\cos\alpha + \frac{\partial U}{\partial y}\cos\beta + \frac{\partial U}{\partial z}\cos\gamma. \qquad (2)$$

Таким образом, для нахождения производной от функции U по направлению l надо найти частные производные $\frac{\partial U}{\partial x}$, $\frac{\partial U}{\partial y}$, $\frac{\partial U}{\partial z}$, направляющие косинусы, задающие это направление, и применить формулу (2).

В случае плоского поля $U = U(x, y)$

$$\cos\beta = \cos\left(\frac{\pi}{2} - \alpha\right) = \sin\alpha, \quad \cos\gamma = 0$$

и

$$\frac{\partial U}{\partial l} = \frac{\partial U}{\partial x}\cos\alpha + \frac{\partial U}{\partial y}\sin\alpha, \qquad (3)$$

где α — угол, образованный направлением l с осью Ox.

Пример. Найти производную от функции $U = xyz$ по направлению l от точки (1, 1,1) к точке (2, 2, 2).

Р е ш е н и е. В данном случае

$$\frac{\partial U}{\partial x} = yz, \quad \frac{\partial U}{\partial y} = xz, \quad \frac{\partial U}{\partial z} = xy,$$

$$\cos\alpha = \frac{2-1}{\sqrt{(2-1)^2 + (2-1)^2 + (2-1)^2}} = \frac{1}{\sqrt{3}},$$

$$\cos\beta = \frac{1}{\sqrt{3}}, \quad \cos\gamma = \frac{1}{\sqrt{3}},$$

$$\frac{\partial U}{\partial l} = \frac{1}{\sqrt{3}}(yz + xz + xy).$$

В частности, в точке (1, 1, 1) имеем

$$\frac{\partial U}{\partial l} = \frac{1}{\sqrt{3}}(1 + 1 + 1) = \frac{3}{\sqrt{3}} = \sqrt{3}.$$

Градиент. Пусть дано скалярное поле $U = U(x, y, z)$, причем функция U имеет частные производные $\frac{\partial U}{\partial x}$, $\frac{\partial U}{\partial y}$ и $\frac{\partial U}{\partial z}$.

Вектор $\frac{\partial U}{\partial x}\boldsymbol{i} + \frac{\partial U}{\partial y}\boldsymbol{j} + \frac{\partial U}{\partial z}\boldsymbol{k}$ называется градиентом скалярного поля U (градиентом функции U) и обозначается grad (иногда ***grad***).

Таким образом,

$$\operatorname{grad} U = \frac{\partial U}{\partial x}\boldsymbol{i} + \frac{\partial U}{\partial y}\boldsymbol{j} + \frac{\partial U}{\partial z}\boldsymbol{k},$$

$$|\operatorname{grad} U| = \sqrt{\left(\frac{\partial U}{\partial x}\right)^2 + \left(\frac{\partial U}{\partial y}\right)^2 + \left(\frac{\partial U}{\partial z}\right)^2}.$$

Направляющие косинусы градиента

$$\cos\alpha_1 = \frac{\frac{\partial U}{\partial x}}{|\operatorname{grad} U|}, \quad \cos\beta_1 = \frac{\frac{\partial U}{\partial y}}{|\operatorname{grad} U|}, \quad \cos\gamma_1 = \frac{\frac{\partial U}{\partial z}}{|\operatorname{grad} U|}.$$

Итак, grad U является векторной величиной. Говорят, что скалярным полем U порождается векторное поле градиента U.

Установим связь между градиентом и производной по направлению и выясним физический смысл градиента.

Возьмем в поле $U = U(P)$ точку P (рис. 63) и найдем grad $U(P)$. Проведем через точку P некоторое направление $\boldsymbol{l}$, образующее с осями координат углы α, β, γ.

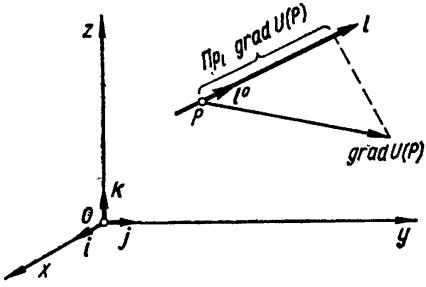

Рис. 63

Пусть $\boldsymbol{l}^\circ = \cos\alpha\, \boldsymbol{i} + \cos\beta \boldsymbol{j} + \cos\gamma \boldsymbol{k}$ — единичный вектор направления $\boldsymbol{l}$. Скалярное произведение векторов $\boldsymbol{l}^\circ$ и grad $U(P)$ имеет вид

$$\boldsymbol{l}^\circ \operatorname{grad} U(P) = \cos\alpha \frac{\partial U}{\partial x} + \cos\beta \frac{\partial U}{\partial y} + \cos\gamma \frac{\partial U}{\partial z}.$$

Известно, что скалярное произведение некоторого вектора на какой-нибудь единичный вектор равно проекции данного вектора на этот единичный вектор.

Таким образом, левая часть полученного равенства представляет собой $\operatorname{Пр}_{\boldsymbol{l}^\circ} \operatorname{grad} U(P)$.

Правая же часть равна $\frac{\partial u}{\partial l}$, т. е.

$$\operatorname{Пр}_{\boldsymbol{l}} \operatorname{grad} U = \frac{\partial U}{\partial l}$$

(так как $\operatorname{Пр}_{\boldsymbol{l}^\circ} \operatorname{grad} U = \operatorname{Пр}_{\boldsymbol{l}} \operatorname{grad} U$).

При изменении направления $\boldsymbol{l}$ изменяется и $\operatorname{Пр}_{\boldsymbol{l}} \operatorname{grad} U$. Очевидно, эта проекция будет иметь наибольшее положительное значение в том случае, когда $\boldsymbol{l}$ совпадает с вектором grad $U(P)$.

Учитывая физический смысл производной по направлению и формулу

$$\frac{\partial U}{\partial l} = \operatorname{Пр}_{\boldsymbol{l}} \operatorname{grad} U,$$

убеждаемся в том, что вектор grad $U(P)$ по величине и направлению есть наибольшая скорость возрастания функции $U = U(P)$ в данной точке P. В этом и состоит физический смысл градиента.

На указанном свойстве градиента основано его широкое применение как в математике, так и в других науках.

В частности, в электрическом поле градиент электрического потенциала grad V есть вектор, равный по величине наибольшему положительному приращению потенциала V, отнесенному к единице

длины, и совпадающий с направлением этого наибольшего приращения (т. е. наибольшая скорость возрастания потенциала).

В случае плоского поля $U = U(x, y)$

$$\operatorname{grad} U = \frac{\partial U}{\partial x} i + \frac{\partial U}{\partial y} j.$$

Покажем, что вектор grad U (P) направлен по нормали к поверхности уровня скалярного поля U (P), проходящей через точку P. Уравнение поверхности уровня

$$U(x, y, z) = C.$$

Уравнения нормали к поверхности уровня имеют вид

$$\frac{X - x}{\frac{\partial U}{\partial x}} = \frac{Y - y}{\frac{\partial U}{\partial y}} = \frac{Z - z}{\frac{\partial U}{\partial z}}$$

(здесь X, Y, Z — координаты текущей точки нормали, x, y, z — координаты точки поверхности, в которой проведена нормаль).

Замечаем, что проекции направляющего вектора нормали к поверхности уровня в точке P (x, y, z) являются проекциями вектора grad U (P). Следовательно, вектор grad U (P) направлен по нормали к этой поверхности. Отсюда следует, например, что и градиент электрического потенциала всюду направлен по нормали (внутренней) к поверхности равного потенциала — к эквипотенциальной поверхности.

В случае плоского поля $U = U(x, y)$ вектор grad U (P) направлен по нормали к линии уровня этого поля, проходящей через точку P (x, y).

Пример. Найти наибольшую скорость возрастания функции $U = x^2 + y^2 + z^2$ в точке P (1, 1, 1).

Р е ш е н и е. Найдем:

$$\operatorname{grad} U = 2xi + 2yj + 2zk,$$

$$\operatorname{grad} U(1, 1, 1) = 2i + 2j + 2k,$$

$$|\operatorname{grad} U(1, 1, 1)| = \sqrt{2^2 + 2^2 + 2^2} = 2\sqrt{3}.$$

Следовательно,

$$v_{\max} = 2\sqrt{3}.$$

Поверхность уровня поля U, проходящая через точку P (1, 1, 1),— сфера $x^2 + y^2 + z^2 = 3$.

Наибольшая скорость возрастания функции U будет в направлении радиуса этой сферы, проходящего через данную точку P (1, 1, 1).

§ 23. ВЕКТОРНОЕ ПОЛЕ. ВЕКТОРНЫЕ ЛИНИИ ПОЛЯ

Понятие векторного поля приведено в § 21. Отметим, что поле вектора a (P) будет задано, если известны три скалярные функции $a_x(P)$, $a_y(P)$, $a_z(P)$ — проекции вектора a (P) на оси координат. Тогда

$$a(P) = a_x(P)\, i + a_y(P)\, j + a_z(P)\, k.$$

Изучение векторного поля удобно начинать с понятия векторных линий поля.

Векторной линией поля $\boldsymbol{a}(P)$ *называется линия, касательная к которой в каждой ее точке совпадает с вектором* $\boldsymbol{a}(P)$*, задающим поле в этой точке.*

Примерами векторных линий поля в электротехнике являются линии вектора напряженности магнитного поля $\boldsymbol{H}$ и линии вектора напряженности электрического поля $\boldsymbol{E}$. Линии вектора $\boldsymbol{E}$ называют силовыми линиями вследствие того, что напряженность электрического поля численно равна силе, действующей в данной точке поля на единичный положительный заряд. Направления вектора этой силы и вектора $\boldsymbol{E}$ всегда совпадают.

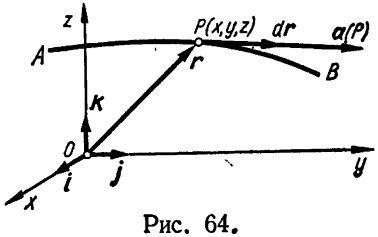

Рис. 64.

Выведем дифференциальные уравнения векторных линий.

Пусть на рис. 64 AB — векторная линия, $\boldsymbol{r}$ — радиус-вектор точки $P(x, y, z)$. Тогда

$$\boldsymbol{r} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}.$$

Найдем

$$d\boldsymbol{r} = dx\boldsymbol{i} + dy\boldsymbol{j} + dz\boldsymbol{k}.$$

Из дифференциального исчисления известно, что вектор $d\boldsymbol{r}$ направлен по касательной к кривой AB в точке P. Следовательно, $d\boldsymbol{r} \parallel \boldsymbol{a}(P)$, но тогда

$$\frac{dx}{a_x} = \frac{dy}{a_y} = \frac{dz}{a_z}. \tag{4}$$

Эти уравнения являются *дифференциальными уравнениями векторных линий*. Если векторное поле по природе поле силы, то векторные линии такого поля будут силовыми линиями.

Примеры

1. Постоянный электрический ток I течет по бесконечно длинному проводу, совпадающему с осью Oz, в направлении снизу вверх (рис. 65). Найти вектор напряженности магнитного поля, создаваемого этим током, в произвольной точке $P(x, y, z)$.

Решение. Рассмотрим достаточно малый элемент провода MN. Согласно закону Био-Савара-Лапласа, вектор напряженности магнитного поля $d\boldsymbol{H}$, создаваемого элементом MN в точке $P(x, y, z)$, находится по формуле (здесь и далее символом $d\boldsymbol{l}$ будем обозначать элемент длины дуги):

$$d\boldsymbol{H} = \frac{1}{4\pi r_1^3}(d\boldsymbol{l} \times \boldsymbol{r}_1),$$

где

$$\boldsymbol{r}_1 = \overrightarrow{MP},\ d\boldsymbol{l} = \overrightarrow{MN}.$$

Очевидно, $d\boldsymbol{H}$ перпендикулярен к плоскости треугольника OPM.

Для определения вектора напряженности магнитного поля H в точке $P(x, y, z)$ надо сложить все векторы dH, возникающие от различных элементов тока Idl, и перейти к пределу, когда длина каждого элемента MN стремится к нулю. Тогда найдем, что

$$H = \int_{-\infty}^{+\infty} \frac{I}{4\pi r_1^3} (dl \times r_1).$$

Если $r = \overrightarrow{OP} = xi + yj + zk$ — радиус-вектор точки $P(x, y, z)$, $\xi = OM$ — расстояние точки M от начала координат, то

$$\overrightarrow{OM} = \xi k; \quad dl = d\xi \cdot k;$$

$$r_1 = \overrightarrow{OP} - \overrightarrow{OM} = xi + yj + (z - \xi) k;$$

$$r_1 = \sqrt{x^2 + y^2 + (z - \xi)^2} = \sqrt{\rho^2 + (z - \xi)^2},$$

где $\rho = \sqrt{x^2 + y^2}$ — расстояние точки $P(x, y, z)$ от оси Oz.

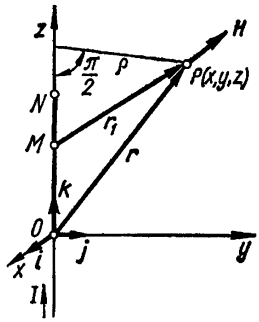

Рис. 65

Найдем

$$dl \times r_1 = \begin{vmatrix} i & j & k \\ 0 & 0 & d\xi \\ x & y & z - \xi \end{vmatrix} = -y d\xi i + x d\xi j.$$

Тогда

$$H = \frac{I}{4\pi} \int_{-\infty}^{+\infty} \frac{d\xi}{\sqrt{(\rho^2 + (z - \xi)^2)^3}} (-yi + xj).$$

Для вычисления интеграла применим подстановку

$$\xi - z = \rho \operatorname{tg} t.$$

Тогда

$$d\xi = \frac{\rho}{\cos^2 t} dt, \text{ при } \xi = \pm\infty, \quad t = \pm\frac{\pi}{2},$$

а потому

$$\int_{-\infty}^{+\infty} \frac{d\xi}{\sqrt{(\rho^2 + (z - \xi)^2)^3}} = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\rho dt}{\cos^2 t \sqrt{(\rho^2 + \rho^2 \operatorname{tg}^2 t)^3}} =$$

$$= \frac{1}{\rho^2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{dt}{\cos^2 t \sec^3 t} = \frac{1}{\rho^2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos t \, dt = \frac{2}{\rho^2}.$$

Следовательно, вектор напряженности магнитного поля H в точке $P(x, y, z)$, создаваемого электрическим током I, текущим по бесконечному проводу,

совпадающему с осью Oz, определяется формулой

$$\boldsymbol{H} = \frac{I}{2\pi\rho^2}(-y\boldsymbol{i} + x\boldsymbol{j}).$$

Очевидно,

$$|\boldsymbol{H}| = \frac{I}{2\pi\rho}.$$

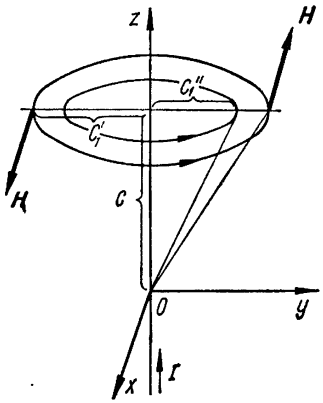

Рис 66

Вектор $\boldsymbol{H}$ направлен перпендикулярно к плоскости, проходящей через ось Oz и точку $P(x, y, z)$ за плоскость чертежа (в сторону, определяемую вращением головки правого винта, ось которого движется поступательно в направлении тока).

2. Найти векторные линии магнитного поля (линии вектора напряженности магнитного поля), образованного постоянным электрическим током I, текущим по бесконечно длинному проводу, совпадающему с осью Oz (рис. 65).

Р е ш е н и е. В предыдущем примере был найден вектор напряженности этого магнитного поля:

$$\boldsymbol{H} = \frac{I}{2\pi\rho^2}(-y\boldsymbol{i} + x\boldsymbol{j}).$$

Проекции вектора $\boldsymbol{H}$ на оси координат:

$$H_x = -\frac{I}{2\pi\rho^2}y,\quad H_y = \frac{I}{2\pi\rho^2}x,\quad H_z = 0.$$

Дифференциальные уравнения векторных линий поля после сокращения на $2\pi\rho^2$:

$$\frac{dx}{-y} = \frac{dy}{x} = \frac{dz}{0}.$$

Последнее отношение $\frac{dz}{0}$ имеет смысл лишь в случае, когда $z = C$ — постоянная величина.

Уравнение $\frac{dx}{-y} = \frac{dy}{x}$ запишем так:

$$x\,dx = -y\,dy.$$

Интегрируя, получим

$$x^2 + y^2 = C_1.$$

Следовательно, векторные линии данного магнитного поля определяются уравнениями

$$\begin{cases} x^2 + y^2 = C_1 \\ z = C. \end{cases}$$

Они представляют собой окружности с центрами на оси Oz, лежащие в плоскостях, перпендикулярных к этой оси (рис. 66).

§ 24. ПОТОК ВЕКТОРА

Пусть дано векторное поле $\boldsymbol{a} = \boldsymbol{a}(P)$. Возьмем в этом поле некоторую поверхность S (рис. 67). Пусть $\boldsymbol{n}$ — нормаль к поверхности S в точке P, $a_n(P) = \text{Пр}_n\,\boldsymbol{a}(P)$ — проекция вектора $\boldsymbol{a}(P)$ на нормаль.

Вычислим величину $K = \iint\limits_S a_n\,(P)\,dS$. Эта величина называется потоком вектора $\boldsymbol{a}\,(P)$ через поверхность S (в выбранном направлении нормалей к S).

Предположим, что поле $\boldsymbol{a}\,(P)$ есть поле скоростей в установившемся потоке несжимаемой жидкости (т. е. в таком потоке, где скорость частиц жидкости в данной точке P не зависит от времени, а зависит лишь от положения точки P). В этом случае поток вектора представляет собой следующее.

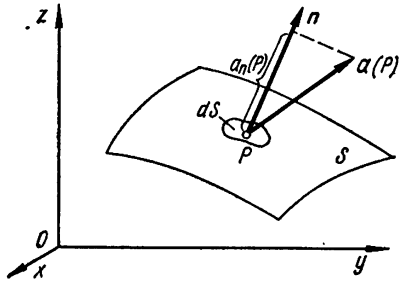

Рис. 67

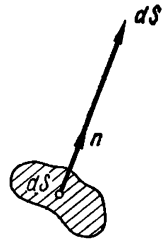

Рис. 68

Выражение $a_n\,(P)\,dS$ равно объему жидкости, протекающей через элемент dS поверхности S за единицу времени. Поэтому поток $K = \iint\limits_S a_n dS$ равен объему жидкости, которая протекает через поверхность S (в указанном направлении нормалей к S) за единицу времени. Таким истолкованием потока вектора $\boldsymbol{a}\,(P)$ удобно пользоваться, если поле $\boldsymbol{a}\,(P)$ имеет и другой физический смысл. Условно считают, что в этом случае $\boldsymbol{a}\,(P)$ есть поле скоростей фиктивного стационарного (установившегося) потока несжимаемой жидкости.

Иногда малую площадку dS представляют в виде вектора $d\boldsymbol{s}$, направленного по нормали к этой площадке, причем $|\,d\boldsymbol{s}\,| = dS$, (рис. 68).

Учитывая это, поток вектора $\boldsymbol{a}$ можно записать так:

$$K = \iint\limits_S \boldsymbol{a} \cdot d\boldsymbol{s} = \int\limits_S \boldsymbol{a} \cdot d\boldsymbol{s}.$$

Так как

$$\boldsymbol{a} = a_x\boldsymbol{i} + a_y\boldsymbol{j} + a_z\boldsymbol{k},$$

$$d\boldsymbol{s} = dydz\boldsymbol{i} + dxdz\boldsymbol{j} + dxdy\boldsymbol{k},$$

то

$$K = \iint\limits_S a_n dS = \iint\limits_S \boldsymbol{a} \cdot d\boldsymbol{s} = \iint\limits_S a_x dydz + a_y dxdz + a_z dxdy.$$

Таким образом, поток вектора $\boldsymbol{a}$ через поверхность S можно записать и в виде поверхностного интеграла по координатам. Сторона поверхности S выбирается при этом в соответствии с ориентацией нормалей к ней.

В электротехнике понятие потока вектора имеет особо важное значение. Так, например, поток вектора плотности тока δ сквозь некоторую поверхность S представляет собой электрический ток i, проходящий через эту поверхность:

$$i = \iint\limits_S \delta ds = \int\limits_S \delta ds.$$

Отсюда видно, что электрический ток i является величиной скалярной, которая не имеет направления, а может принимать только положительные или отрицательные значения, что и учитывается при расчете электрических цепей.

Аналогично магнитный поток Φ представляет собой поток вектора магнитной индукции $\boldsymbol{B}$ сквозь некоторую поверхность S:

$$\Phi = \iint\limits_S \boldsymbol{B} \cdot d\boldsymbol{s} = \int\limits_S \boldsymbol{B} \cdot d\boldsymbol{s},$$

где $\boldsymbol{B} = \mu_a \boldsymbol{H}$ — вектор магнитной индукции, характеризующий магнитное поле в каждой точке; μ_a — абсолютная магнитная проницаемость среды, в которой существует магнитное поле; $\boldsymbol{H}$ — вектор напряженности магнитного поля.

В электрическом поле рассматривается поток вектора напряженности электрического поля $\boldsymbol{E}$ сквозь поверхность S:

$$N = \iint\limits_S \boldsymbol{E} \cdot d\boldsymbol{s},$$

что дает возможность в ряде случаев довольно просто рассчитать исследуемое электрическое поле, т. е. определить $\boldsymbol{E}(x, y, z)$ по заданному распределению электрических зарядов, создающих поле в пространстве. Особый интерес представляет случай, когда поверхность S замкнута и ограничивает некоторый объем V. Тогда поток вектора записывается так:

$$\oint\limits_S a_n \cdot dS, \text{ или } \oint\limits_S \boldsymbol{a} \cdot d\boldsymbol{s}.$$

В этом случае за направление вектора $\boldsymbol{n}$ обычно берут направление внешней нормали. Говорят, что поток вектора через замкнутую поверхность представляет собой разность между числом векторных линий, выходящих и входящих в объем, ограниченный этой поверхностью.

В точках поверхности S, где векторные линии выходят из объема V, внешняя нормаль к поверхности образует с вектором $\boldsymbol{a}$ острый угол, $\boldsymbol{a} \cdot d\boldsymbol{s} > 0$ (жидкость вытекает); в точках поверхности, где векторные линии входят в объем (жидкость втекает в него), $\boldsymbol{a} \cdot d\boldsymbol{s} < 0$. Поэтому поток вектора $\boldsymbol{a}$ через замкнутую поверхность S можно рассматривать, как разность между количествами жидкости, вытекающей из объема V и втекающей в него за единицу времени.

Если жидкость несжимаема, то ее количество внутри объема V должно оставаться неизменным.

Пусть поток $\oint_S a_n dS > 0$. Это значит, что вытекает из V жидкости больше, чем втекает в него, что возможно только тогда, когда внутри V есть дополнительные источники, питающие поток. Если же $\oint_S a_n\, dS < 0$, то внутри V имеются стоки, поглощающие избыток жидкости. Если в объеме V нет ни источников, ни стоков, то $\oint_S a_n dS = 0$. (Если $\oint_S a_n dS = 0$, то из этого еще не следует, что внутри объема V нет ни источников, ни стоков, так как они могут быть, но действие их взаимно компенсируется).

Можно сказать, что источники — это точки, где векторные линии начинаются, а стоки — точки, где векторные линии кончаются. В электрическом поле неподвижных и неизменяющихся по величине зарядов (в электростатическом поле), согласно теореме Гаусса, поток вектора напряженности электрического поля $\boldsymbol{E}$ сквозь любую замкнутую поверхность, взятую в однородной и изотропной среде с абсолютной диэлектрической проницаемостью ε_a, равен отношению количества электричества Q, заключенного в объеме V, ограниченном этой поверхностью, к величине ε_a:

$$\oint_S \boldsymbol{E} \cdot d\boldsymbol{s} = \frac{Q}{\varepsilon_a}.$$

Если $Q > 0$, то $\oint_S \boldsymbol{E} \cdot d\boldsymbol{s} > 0$; если же $Q < 0$, то и $\oint_S \boldsymbol{E} \cdot d\boldsymbol{s} < 0$. Отсюда можно заключить, что в электростатическом поле положительные заряды являются источниками (истоками) силовых линий, а отрицательные заряды — стоками.

В магнитном поле справедлив принцип непрерывности магнитного потока, согласно которому магнитный поток сквозь любую замкнутую поверхность равен нулю:

$$\Phi = \oint_S \boldsymbol{B} \cdot d\boldsymbol{s} = 0.$$

Для магнитного поля это физически означает, что в нем нет точек, где линии вектора магнитной индукции начинаются (нет истоков), и нет точек, где они кончаются (нет стоков). Линии вектора $\boldsymbol{B}$ замкнуты (непрерывны). Сколько их входит внутрь объема V, ограниченного замкнутой поверхностью S, столько же из этого объема их и выходит. Единственной причиной, вызывающей образование магнитного поля, является электрический ток.

В свою очередь, поток вектора плотности полного тока $\boldsymbol{\delta}$ сквозь любую замкнутую поверхность равен нулю:

$$\oint_S \boldsymbol{\delta} \cdot d\boldsymbol{s} = 0.$$

Это уравнение выражает фундаментальный принцип теории электромагнитного поля — принцип непрерывности электрического тока (или принцип замкнутости линий тока). Линии вектора магнитной индукции и линии вектора плотности полного тока замкнуты и сцеплены между собой (охватывают друг друга).

В математическом анализе имеется формула Остроградского:

$$\oint\limits_S a_x dydz + a_y dxdz + a_z dxdy = \iiint\limits_V \left(\frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z}\right) dv.$$

Как мы теперь уже знаем, в левой части формулы записан поток вектора $\boldsymbol{a}$ через замкнутую поверхность S (в направлении внешней нормали к S). Эту формулу удобно использовать при вычислении потока вектора через замкнутую поверхность. Однако ее можно применить и в том случае, когда поверхность S не является замкнутой, но тогда надо дополнить заданную поверхность S до замкнутой некоторой поверхностью S_1 (желательно наиболее простым способом). Получим

$$\iint\limits_S a_n dS + \iint\limits_{S_1} a_n dS = \oint\limits_{S+S_1} a_n dS = \iiint\limits_V \left(\frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z}\right) dv.$$

Примечание. В теории электромагнитного поля используется теорема Остроградского — Гаусса:

$$\oint\limits_S E \cdot d\boldsymbol{s} = \oint\limits_S E_x dydz + E_y dxdz + E_z dxdy = \iiint\limits_V \left(\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_x}{\partial z}\right) dv,$$

где E_x, E_y, E_z — проекции вектора напряженности электрического поля $\boldsymbol{E}$ на оси прямоугольной декартовой системы координат.

Интересующий нас поток через поверхность S запишется так:

$$\iint\limits_S a_n dS = \oint\limits_{S+S_1} a_n dS - \iint\limits_{S_1} a_n dS.$$

При этом $\oint\limits_{S+S_1} a_n dS$ находится по формуле Остроградского, а $\iint\limits_{S_1} a_n dS$ — обычным способом (при рациональном выборе поверхности S_1 это, как правило, не представляет особого труда).

Пример. Найти поток радиуса-вектора $\boldsymbol{r}$ через часть плоскости $x + y + z = 1$, заключенную в первом октанте; направление нормали к плоскости указано на чертеже (рис. 69).

Р е ш е н и е. *Первый способ.* В данном случае

$$\boldsymbol{r} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}.$$

Нормаль к поверхности S (к плоскости $x + y + z = 1$) можно записать так:

$$\boldsymbol{n} = \boldsymbol{i} + \boldsymbol{j} + \boldsymbol{k}.$$

Проекция вектора $\boldsymbol{r}$ на $\boldsymbol{n}$ будет

$$r_n = \frac{\boldsymbol{r} \cdot \boldsymbol{n}}{|\boldsymbol{n}|} = \frac{x + y + z}{\sqrt{3}},$$

$$\iint\limits_S r_n dS = \iint\limits_S \frac{x+y+z}{\sqrt{3}}\, dS = \iint\limits_S \frac{dS}{\sqrt{3}} = \frac{S}{\sqrt{3}},$$

так как на поверхности S имеем $x + y + z = 1$.

Найдем S. Для этого вычислим площадь равностороннего треугольника со стороной $a = \sqrt{2}$:

$$S = \frac{a^2\sqrt{3}}{4} = \frac{2\sqrt{3}}{4} = \frac{\sqrt{3}}{2}.$$

Тогда

$$\iint\limits_S r_n dS = \frac{\sqrt{3}}{2\sqrt{3}} = \frac{1}{2}.$$

Второй способ. Дополним поверхность S до замкнутой (полной поверхности пирамиды) треугольниками S_1, S_2, S_3, лежащими в координатных плоскостях. Тогда по формуле Остроградского

$$\oint\limits_{S+S_1+S_2+S_3} r_n dS = \iiint\limits_V \left(\frac{\partial r_x}{\partial x} + \frac{\partial r_y}{\partial y} + \frac{\partial r_z}{\partial z}\right) dv =$$

$$= \iiint\limits_V (1+1+1)\, dv = 3 \iiint\limits_V dv = 3v = 3 \cdot \frac{1}{3} S_1 \cdot 1 = \frac{1}{2};$$

$$\iint\limits_{S_1} r_n dS = \iint\limits_{S_1} 0 \cdot dS = 0,$$

так как для любой точки $(x, y, 0)$ треугольника S_1 радиус-вектор $\boldsymbol{r} = x\boldsymbol{i} + y\boldsymbol{j}$ лежит в плоскости этого треугольника, а нормаль (внешняя) к этому треугольнику направлена по оси Oz (вниз), следовательно, $r_n \equiv 0$ на треугольнике S_1. Аналогично

$$\iint\limits_{S_2} r_n \cdot dS = 0, \qquad \iint\limits_{S_3} r_n \cdot dS = 0.$$

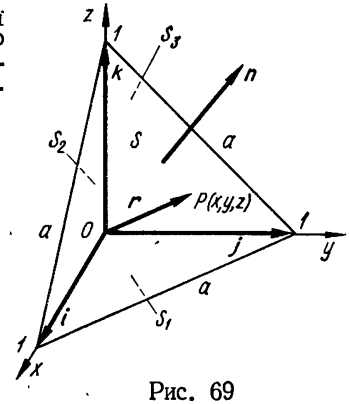

Рис. 69

Тогда $\iint\limits_S r_n dS = \oint\limits_{S+S_1+S_2+S_3} r_n dS - \iint\limits_{S_1} r_n dS -$

$$- \iint\limits_{S_2} r_n dS - \iint\limits_{S_3} r_n dS =$$

$$= \frac{1}{2} - 0 - 0 - 0 = \frac{1}{2}.$$

§ 25. РАСХОДИМОСТЬ ПОЛЯ. ФОРМУЛА ОСТРОГРАДСКОГО

Пусть дано векторное поле $\boldsymbol{a}(P)$. Возьмем в этом поле некоторую точку P и поместим ее внутрь небольшой замкнутой поверхности S, ограничивающей объем Δv. Найдем $\oint\limits_S a_n dS$ — поток вектора $\boldsymbol{a}$

через замкнутую поверхность S (в направлении внешней нормали к S).

Вычислим величину

$$\frac{\oint\limits_S a_n dS}{\Delta v}.$$

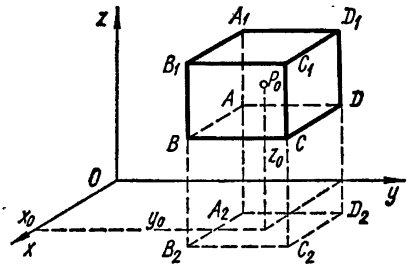

Рис. 70

Предположим, что существует предел этой величины при условии, что поверхность S стягивается в точку P. Этот предел называется расходимостью векторного поля $\boldsymbol{a}$ в точке P (дивергенцией вектора $\boldsymbol{a}$) и обозначается: $\operatorname{div} \boldsymbol{a}$. Таким образом, по определению:

$$\operatorname{div} \boldsymbol{a}(P) = \lim_{\Delta v \to 0} \frac{\oint\limits_S a_n dS}{\Delta v}.$$

Ясно, что $\operatorname{div} \boldsymbol{a}(P)$ есть скалярная величина. Она является важной характеристикой векторного поля. Говорят, что векторное поле $\boldsymbol{a}$ порождает скалярное поле расходимости $\operatorname{div} \boldsymbol{a}$.

Если в точке P $\operatorname{div} \boldsymbol{a}(P) > 0$, то, очевидно, для достаточно малого Δv

$$\frac{\oint\limits_S a_n dS}{\Delta v} > 0,$$

так как $\Delta v > 0$, то и $\oint\limits_S a_n ds > 0$. Учитывая физический смысл потока вектора (см. § 24), заключаем, что в данном случае в точке P находится источник поля.

Если в точке P $\operatorname{div} \boldsymbol{a}(P) < 0$, то в этой точке находится сток.

В точках, где $\operatorname{div} \boldsymbol{a} > 0$, векторные линии начинаются, а в точках, где $\operatorname{div} \boldsymbol{a} < 0$, векторные линии кончаются.

В теории электромагнитного поля дивергенция вектора электрического смещения равна объемной плотности электрического заряда в данной точке ($\operatorname{div} \boldsymbol{D} = \rho$). В тех точках, где $\rho > 0$ (расположен положительный заряд), линии вектора $\boldsymbol{D}$ начинаются — имеется исток. В тех же точках, где $\rho < 0$ (отрицательный заряд), линии вектора $\boldsymbol{D}$ заканчиваются — имеется сток.

Численное значение расходимости характеризует интенсивность источника или стока. (Под интенсивностью источника в физике часто понимают количество энергии, излучаемой источником, отнесенной к единице телесного угла).

Если рассматривать векторное поле как поле скоростей частиц текущей жидкости, то дивергенция характеризует обильность ис-

точника, находящегося в сколь угодно малом объеме, окружающем точку P.

Векторные поля, у которых в каждой точке $\operatorname{div} \boldsymbol{a} = 0$, называются соленоидальными. В таком поле нет источников и стоков, векторные линии не могут на конечном расстоянии нигде ни начинаться, ни кончаться, они могут уходить в бесконечность, быть замкнутыми или же начинаться и кончаться у границ поля.

Примерами соленоидальных полей являются: магнитное поле, поле вектора плотности полного тока, поле скоростей частиц вращающейся жидкости. Векторные линии этих полей — замкнутые линии. Выразим расходимость вектора через его проекции.

Пусть вектор $\boldsymbol{a} = a_x\boldsymbol{i} + a_y\boldsymbol{j} + a_z\boldsymbol{k}$. Найдем $\operatorname{div} \boldsymbol{a}$ в точке $P_0(x_0, y_0, z_0)$. Возьмем прямоугольный параллелепипед с центром в точке P_0 и размерами $2\Delta x$, $2\Delta y$, $2\Delta z$. Пусть грани параллелепипеда параллельны соответствующим координатным плоскостям (рис. 70): $|AB| = 2\Delta x$, $|BC| = 2\Delta y$, $|AA_1| = 2\Delta z$. Поток вектора через поверхность этого параллелепипеда равен:

$$\oint_S a_n dS = \iint_{S_1} a_n dS + \iint_{S_2} a_n dS + \cdots + \iint_{S_6} a_n dS,$$

где S_1 — грань ADD_1A_1, S_2 — грань BCC_1B_1 и т. д.

Рассмотрим интегралы по противоположным параллельным граням S_1 и S_2. На грани S_2 внешняя нормаль совпадает с осью Ox, поэтому $a_n = a_x$. Очевидно, на грани S_1 выполняется условие: $a_n = -a_x$. При этом на грани S_2 имеем $x = x_0 + \Delta x$, на грани S_1 соответственно $x = x_0 - \Delta x$. На каждой из этих граней y меняется от $y_0 - \Delta y$ до $y_0 + \Delta y$; z меняется от $z_0 - \Delta z$ до $z_0 + \Delta z$.

Следовательно,

$$\iint_{S_1} a_n dS + \iint_{S_2} a_n dS = \iint_{S_2} a_x dS - \iint_{S_1} a_x dS =$$

$$= \int_{y_0-\Delta y}^{y_0+\Delta y} \int_{z_0-\Delta z}^{z_0+\Delta z} a_x(x_0 + \Delta x, y, z)\, dydz -$$

$$- \int_{y_0-\Delta y}^{y_0+\Delta y} \int_{z_0-\Delta z}^{z_0+\Delta z} a_x(x_0 - \Delta x, y, z)\, dydz =$$

$$= \int_{y_0-\Delta y}^{y_0+\Delta y} \int_{z_0-\Delta z}^{z_0+\Delta z} (a_x(x_0 + \Delta x, y, z) - a_x(x_0 - \Delta x, y, z))\, dydz.$$

К разности в подынтегральном выражении применим формулу Лагранжа по переменной x, а после этого используем теорему о среднем для интеграла. Получим

$$\iint_{S_1} a_n dS + \iint_{S_2} a_n dS =$$

$$= \int_{y_0-\Delta y}^{y_0+\Delta y} \int_{z_0-\Delta z}^{z_0+\Delta z} \frac{\partial}{\partial x} a_x (x_0 + \theta \Delta x, y, z)\, 2\Delta x dy dz =$$

$$= \left(\frac{\partial a_x}{\partial x}\right)_1 \cdot 2\Delta x \cdot 2\Delta y \cdot 2\Delta z.$$

Здесь $|\theta| < 1$, $\left(\frac{\partial a_x}{\partial x}\right)_1$ — значение производной $\frac{\partial a_x}{\partial x}$ в некоторой точке внутри параллелепипеда.

Аналогично вычислим поток через остальные две пары параллельных граней параллелепипеда. Поэтому

$$\oint_S a_n dS = \left(\left(\frac{\partial a_x}{\partial x_1}\right)_1 + \left(\frac{\partial a_y}{\partial y}\right)_2 + \left(\frac{\partial a_z}{\partial z}\right)_3\right) \cdot 8\Delta x \Delta y \Delta z.$$

Здесь индексы 1, 2, 3 означают, что соответствующие частные производные находятся в некоторых внутренних точках параллелепипеда. Заметим, что $\Delta v = 8\Delta x \Delta y \Delta z$. Тогда

$$\operatorname{div} \boldsymbol{a}\,(P_0) = \lim_{\Delta v \to 0} \frac{\oint_S a_n dS}{\Delta v} = \lim_{\Delta v \to 0} \left(\left(\frac{\partial a_x}{\partial x}\right)_1 + \left(\frac{\partial a_y}{\partial y}\right)_2 + \left(\frac{\partial a_z}{\partial z}\right)_3\right) =$$

$$= \frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z} \text{ в точке } P_0 (x_0, y_0, z_0).$$

Следовательно, в точке $P\ (x, y, z)$

$$\operatorname{div} \boldsymbol{a} = \frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z}.$$

Мы получили формулу для вычисления расходимости векторного поля $\boldsymbol{a}$, которую легко применять, если известны проекции вектора на оси координат.

Примеры.

1. Найти $\operatorname{div} \boldsymbol{a}$, если $\boldsymbol{a} = x^2\boldsymbol{i} + y^2\boldsymbol{j} + z^2\boldsymbol{k}$. В данном случае

$$a_x = x^2,\ a_y = y^2,\ a_y = z^2;$$

$$\frac{\partial a_x}{\partial x} = 2x,\ \frac{\partial a_y}{\partial y} = 2y,\ \frac{\partial a_z}{\partial z} = 2z;$$

$$\operatorname{div} \boldsymbol{a} = 2\,(x + y + z).$$

2. Найти расходимость радиуса-вектора $\boldsymbol{r}$ точки $P\ (x, y, z)$.

Р е ш е н и е. $\boldsymbol{r} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$; $r_x = x$, $r_y = y$, $r_z = z$;

$$\frac{\partial r_x}{\partial x} = \frac{\partial r_y}{\partial y} = \frac{\partial r_z}{\partial z} = 1.$$

Поэтому

$$\operatorname{div} \boldsymbol{r} = \frac{\partial r_x}{\partial x} + \frac{\partial r_y}{\partial y} + \frac{\partial r_z}{\partial z} = 3.$$

Отметим некоторые свойства расходимости:

1. Если $\boldsymbol{a}$ — постоянный вектор, то $\operatorname{div} \boldsymbol{a} = 0$.
2. $\operatorname{div}(\boldsymbol{a} \pm \boldsymbol{b}) = \operatorname{div} \boldsymbol{a} \pm \operatorname{div} \boldsymbol{b}$.
3. Пусть U — скалярная функция, $\boldsymbol{a}$ — вектор. Тогда

$$\operatorname{div}(U\boldsymbol{a}) = U \operatorname{div} \boldsymbol{a} + \boldsymbol{a} \operatorname{grad} U.$$

Эти свойства легко проверяются, если применить полученную выше формулу для нахождения расходимости. В качестве упражнения рекомендуем проверить их. Несколько позже будет показано, как это можно сделать иначе (см. § 28).

Запишем теперь известную в анализе формулу Остроградского используя понятия теории поля.

Пусть S — замкнутая поверхность, ограничивающая объем V, a_x, a_y, a_z — функции от x, y, z, непрерывные вместе со своими частными производными. Тогда

$$\oint_S a_x dydz + a_y dxdz + a_z dxdy = \iiint_V \left(\frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z}\right) dv,$$

где поверхностный интеграл вычисляется по внешней стороне поверхности S.

Если функции a_x, a_y, a_z рассматривать как проекции некоторого вектора $\boldsymbol{a}$, то левая часть формулы Остроградского равна потоку вектора $\boldsymbol{a}$ через замкнутую поверхность S (в направлении внешней нормали к S), а

$$\frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z} = \operatorname{div} \boldsymbol{a},$$

и формула Остроградского примет вид

$$\oint_S a_n dS = \iiint_V \operatorname{div} \boldsymbol{a} dv.$$

Она выражает следующий важный результат: *поток векторного поля через некоторую замкнутую поверхность* (в направлении внешней нормали к ней) *равен тройному интегралу от расходимости этого поля, вычисленному по объему, ограниченному данной поверхностью.*

В теории электромагнитного поля теорема Гаусса — Остроградского (см. § 24) обычно записывается так:

$$\oint_S \boldsymbol{E} d\boldsymbol{s} = \iiint_V \operatorname{div} \boldsymbol{E} dv.$$

§ 26. ЦИРКУЛЯЦИЯ ВЕКТОРА

Возьмем поле вектора $\boldsymbol{a} = a_x\boldsymbol{i} + a_y\boldsymbol{j} + a_z\boldsymbol{k}$ и некоторую кривую AB (рис. 71). Пусть

$$a_\tau = \operatorname{Пр}_\tau \boldsymbol{a},$$

где τ — единичный вектор касательной к кривой AB в точке P. Рассмотрим криволинейный интеграл $\int\limits_{AB} a_\tau dl$. Этот интеграл называется циркуляцией вектора a вдоль кривой AB.

Если поле a — силовое, то $\int\limits_{AB} a_\tau dl$ имеет простой физический смысл: это работа сил поля вдоль кривой AB.

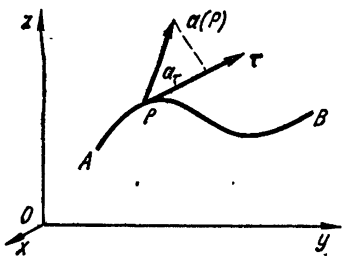

Рис. 71

В частности, циркуляция вектора напряженности электрического поля E вдоль кривой AB, т. е. $\int\limits_{AB} E_\tau dl$, представляет собой работу, совершаемую силами электрического поля по перемещению единичного положительного заряда из точки A в точку B. Напомним, что вектор напряженности электрического поля E является силовой характеристикой поля.

Очевидно,

$$\int\limits_{AB} a_\tau dl = -\int\limits_{BA} a_\tau dl.$$

Во многих вопросах рассматривают циркуляцию вектора вдоль замкнутого контура L

$$\oint\limits_L a_\tau dl.$$

Пусть r — радиус-вектор точки P на контуре L. Известно, что вектор dr направлен по касательной к контуру в точке P, а $|dr| = dl$. Поэтому удобно рассматривать вектор $dl = dr$ и записывать циркуляцию вектора a вдоль контура L в виде

$$\oint\limits_L a dr, \text{ или } \oint\limits_L a dl.$$

Так как

$$r = xi + yj + zk,$$

то

$$dr = dx \cdot i + dy \cdot j + dz \cdot k,$$

тогда

$$\oint\limits_L a_\tau dl = \oint\limits_L a dr = \oint\limits_L a_x dx + a_y dy + a_z dz.$$

Способы вычисления криволинейных интегралов подробно изучаются в соответствующей теме курса математического анализа, поэтому мы здесь на них не останавливаемся. Отметим лишь, что вдоль замкнутых векторных линий циркуляция не равна нулю, ибо

в каждой точке векторной линии скалярное произведение $\boldsymbol{a} \cdot d\boldsymbol{r}$ сохраняет один и тот же знак:

$$\boldsymbol{a} \cdot d\boldsymbol{r} = \begin{cases} adr, & \text{если направление вектора } \boldsymbol{a} \text{ на векторной линии совпадает с направлением ее обхода} \\ -adr & \text{— в противном случае.} \end{cases}$$

Следовательно,

$$\oint_L \boldsymbol{a} \cdot d\boldsymbol{r} = \pm \oint_L adr \neq 0.$$

Так, в частности, для магнитного поля циркуляция вектора магнитной индукции $\boldsymbol{B}$ вдоль любой линии вектора $\boldsymbol{B}$ всегда отлична от нуля

$$\oint_L \boldsymbol{B} dl \neq 0,$$

так как магнитное поле соленоидально ($\operatorname{div} \boldsymbol{B} = 0$), и линии вектора $\boldsymbol{B}$ замкнуты.

§ 27. ВИХРЬ (РОТОР) ВЕКТОРА. ФОРМУЛА СТОКСА

Возьмем в поле вектора $\boldsymbol{a}$ точку P_0. Пусть ΔS — достаточно малая поверхность, содержащая точку P_0, L — контур этой поверхности, $\boldsymbol{n}$ — нормаль к ΔS в точке P_0. Будем считать, что направление нормали $\boldsymbol{n}$ и положительный обход контура L согласованы (рис. 72). Вычислим циркуляцию вектора $\boldsymbol{a}$ вдоль контура L:

$$\oint_L a_\tau dl = \oint_L adr.$$

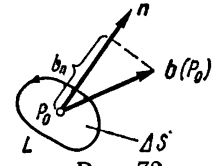

Рис. 72

Предположим, что существует

$$\lim_{\Delta S \to 0} \frac{\oint_L a_\tau dl}{\Delta S} = b_n \tag{5}$$

(предел находится при условии, что поверхность ΔS стягивается в точку P_0).

Величину этого предела обозначили через b_n, учитывая, что она, вообще говоря, зависит от положения контура L, т. е. от направления нормали $\boldsymbol{n}$ к поверхности ΔS.

Рассмотрим вектор $\boldsymbol{b}$ такой, что

$$\Pi p_{\boldsymbol{n}} \boldsymbol{b} = b_n, \tag{6}$$

где b_n находится по формуле (5). Формула (6) вполне определяет вектор $\boldsymbol{b}$ в данной точке поля P_0, так как известна проекция этого вектора на любое направление $\boldsymbol{n}$. Этот вектор $\boldsymbol{b}$ называется в и х р е м, или р о т о р о м данного векторного поля $\boldsymbol{a}$ (вихрем, или ротором вектора $\boldsymbol{a}$) в указанной точке P_0. Обозначается вихрь (ротор) так:

$$\operatorname{rot} \boldsymbol{a}$$

(читается: «ротор вектора»). Таким образом,

$$\operatorname{rot}_n \boldsymbol{a} = \lim_{\Delta S \to 0} \frac{\oint\limits_L a_\tau dl}{\Delta S}, \qquad (7)$$

где

$$\operatorname{rot}_n \boldsymbol{a} = \Pi p_n \operatorname{rot} \boldsymbol{a}.$$

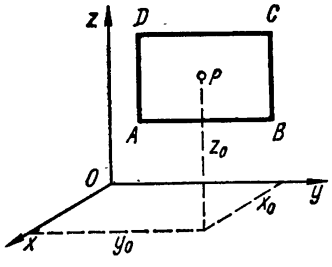

Рис. 73

Говорят, что данное векторное поле $\boldsymbol{a}$ порождает другое векторное поле — поле $\operatorname{rot} \boldsymbol{a}$.

Вихрь векторного поля $\boldsymbol{a}$ является очень важной характеристикой этого поля. Отметим, прежде всего, что вектор $\operatorname{rot} \boldsymbol{a}$ в данной точке P поля $\boldsymbol{a}$ обладает следующим свойством: если нормаль $\boldsymbol{n}$ к данной площадке ΔS совпадает с направлением $\operatorname{rot} \boldsymbol{a}$, то циркуляция $\oint\limits_L a_\tau \, dl$ вектора $\boldsymbol{a}$ вдоль контура этой площадки имеет наибольшее значение по сравнению с циркуляцией при любом другом положении этой площадки (когда нормаль к ней не совпадает с направлением $\operatorname{rot} \boldsymbol{a}$ в данной точке поля $\boldsymbol{a}$).

Действительно, из формулы (7) на основании теоремы о связи между переменной величиной, ее пределом и бесконечно малой величиной следует, что

$$\oint_L a_\tau dl = \operatorname{rot}_n \boldsymbol{a} \cdot \Delta S + \varepsilon \Delta S,$$

где $\varepsilon \to 0$ при $\Delta S \to 0$.

Тогда

$$\oint_L a_\tau dl \approx \Delta S \cdot \operatorname{rot}_n \boldsymbol{a},$$

но величина $\operatorname{rot}_n \boldsymbol{a}$ имеет наибольшее значение, если $\boldsymbol{n}$ совпадает с направлением $\operatorname{rot} \boldsymbol{a}$, следовательно, $\oint\limits_L a_\tau dl$ в этом случае также имеет наибольшее значение. Если рассматривать векторное поле как поле скоростей частиц текущей жидкости, то ротор (вихрь) характеризует угловую скорость вращения сколь угодно малого объема, окружающего точку P.

Понятие вихря поля $\boldsymbol{a}$ введено независимо от того, в какой системе координат заданы проекции вектора $\boldsymbol{a}$. Однако для решения задач, где приходится применять операцию нахождения вихря, целесообразно уметь определять проекции вектора $\operatorname{rot} \boldsymbol{a}$, например, на оси декартовой системы координат.

Чтобы найти проекцию вихря на ось Ox, в качестве площадки ΔS возьмем прямоугольник $ABCD$ с центром в точке $P(x_0, y_0, z_0)$, параллельный плоскости yOz (рис. 73).

Пусть $|AB| = 2\Delta y$, $|AD| = 2\Delta z$. На стороне AB: $x = x_0$, $z = z_0 - \Delta z$, y изменяется от $y_0 - \Delta y_0$ до $y_0 + \Delta y$, $a_\tau = a_y$. На стороне CD: $x = x_0$; $z = z_0 + \Delta z$, y изменяется от $y_0 + \Delta y$ до $y_0 - \Delta y$, $a_\tau = -a_y$.

Тогда

$$\int_{AB} a_\tau dl + \int_{CD} a_\tau dl = \int_{y_0-\Delta y}^{y_0+\Delta y} a_y(x_0, y, z_0 - \Delta z)\, dy + \int_{y_0+\Delta y}^{y_0-\Delta y} (-a_y(x_0, y, z_0 + \Delta z))(-dy) =$$

$$= -\int_{y_0-\Delta y}^{y_0+\Delta y} (a_y(x_0, y, z_0 + \Delta z) - a_y(x_0, y, z_0 - \Delta z))\, dy =$$

$$= -\int_{y_0-\Delta y}^{y_0+\Delta y} \frac{\partial}{\partial z} a_y(x_0, y, z_0 + \theta_1 \Delta z)\, 2\Delta z \cdot dy$$

(к разности в подынтегральном выражении применили формулу Лагранжа по переменной z; $|\theta_1| < 1$). К полученному интегралу применим теорему о среднем. Тогда

$$\int_{AB} a_\tau dl + \int_{CD} a_\tau dl = -4\left(\frac{\partial a_y}{\partial z}\right)_1 \Delta z \Delta y,$$

где $\left(\frac{\partial a_y}{\partial z}\right)_1$ — значение производной $\frac{\partial a_y}{\partial z}$ в некоторой точке прямоугольника $ABCD$. Аналогично найдем, что

$$\int_{BC} a_\tau dl + \int_{DA} a_\tau dl = 4\left(\frac{\partial a_z}{\partial y}\right)_2 \Delta z \Delta y,$$

где $\left(\frac{\partial a_z}{\partial y}\right)_2$ — значение производной $\frac{\partial a_z}{\partial y}$ в некоторой точке прямоугольника $ABCD$.

Тогда

$$\oint_{ABCD} a_\tau dl = \left(\left(\frac{\partial a_z}{\partial y}\right)_2 - \left(\frac{\partial a_y}{\partial z}\right)_1\right) 4\Delta z \Delta y.$$

Так как $\Delta S = 4\Delta z \Delta y$, то

$$\frac{\oint_{ABCD} a_\tau dl}{\Delta S} = \left(\left(\frac{\partial a_z}{\partial y}\right)_2 - \left(\frac{\partial a_y}{\partial z}\right)_1\right).$$

При $\Delta S \to 0$, т. е. при $\Delta y \to 0$ и $\Delta z \to 0$, получим

$$\operatorname{rot}_x \boldsymbol{a} = \frac{\partial a_z}{\partial y} - \frac{\partial a_y}{\partial z}.$$

Принимая в качестве ΔS прямоугольники, перпендикулярные соответственно к осям Oy и Oz, аналогично найдем, что

$$\operatorname{rot}_y \boldsymbol{a} = \frac{\partial a_x}{\partial z} - \frac{\partial a_z}{\partial x}, \quad \operatorname{rot}_z \boldsymbol{a} = \frac{\partial a_y}{\partial x} - \frac{\partial a_x}{\partial y}.$$

Зная проекции вектора $\operatorname{rot} \boldsymbol{a}$ на оси координат, можно найти его модуль и направляющие косинусы. В частности,

$$|\operatorname{rot} \boldsymbol{a}| = \sqrt{\left(\frac{\partial a_z}{\partial y} - \frac{\partial a_y}{\partial z}\right)^2 + \left(\frac{\partial a_x}{\partial z} - \frac{\partial a_z}{\partial x}\right)^2 + \left(\frac{\partial a_y}{\partial x} - \frac{\partial a_x}{\partial y}\right)^2}.$$

Для удобства запоминания вихрь вектора $\boldsymbol{a}$ можно записать в виде такого символического определителя:

$$\operatorname{rot} \boldsymbol{a} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ a_x & a_y & a_z \end{vmatrix},$$

где во второй строке написаны символы для операций дифференцирования по x, y, z.

Примеры

1. Найти вихрь вектора $\boldsymbol{a} = z\boldsymbol{i} + x\boldsymbol{j} + y\boldsymbol{k}$.

Р е ш е н и е. В данном случае $a_x = z$, $a_y = x$, $a_z = y$.

$$\operatorname{rot} \boldsymbol{a} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ z & x & y \end{vmatrix} = \left(\frac{\partial y}{\partial y} - \frac{\partial x}{\partial z}\right)\boldsymbol{i} + \left(\frac{\partial z}{\partial z} - \frac{\partial y}{\partial x}\right)\boldsymbol{j} + \left(\frac{\partial x}{\partial x} - \frac{\partial z}{\partial y}\right)\boldsymbol{k} = \boldsymbol{i} + \boldsymbol{j} + \boldsymbol{k}.$$

2. Найти вихрь вектора $\boldsymbol{r} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$.

Р е ш е н и е.

$$\operatorname{rot} \boldsymbol{r} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ x & y & z \end{vmatrix} = \left(\frac{\partial z}{\partial y} - \frac{\partial y}{\partial z}\right)\boldsymbol{i} + \left(\frac{\partial x}{\partial z} - \frac{\partial z}{\partial x}\right)\boldsymbol{j} + \left(\frac{\partial y}{\partial x} - \frac{\partial x}{\partial y}\right)\boldsymbol{k} = 0 \cdot \boldsymbol{i} + 0 \cdot \boldsymbol{j} + 0 \cdot \boldsymbol{k} = 0.$$

Отметим некоторые свойства вихря.

1. *Если* $\boldsymbol{a}$ *— постоянный вектор, то* $\operatorname{rot} \boldsymbol{a} = 0$.
2. $\operatorname{rot}(\boldsymbol{a} \pm \boldsymbol{b}) = \operatorname{rot} \boldsymbol{a} \pm \operatorname{rot} \boldsymbol{b}$.
3. $\operatorname{rot}(v\boldsymbol{a}) = v \cdot \operatorname{rot} \boldsymbol{a} + \operatorname{grad} v \times \boldsymbol{a}$.

4. $\operatorname{div}(\operatorname{rot} \boldsymbol{a}) = 0$.

5. $\operatorname{div}(\boldsymbol{a} \times \boldsymbol{b}) = \boldsymbol{b} \cdot \operatorname{rot} \boldsymbol{a} - \boldsymbol{a} \cdot \operatorname{rot} \boldsymbol{b}$.

Перечисленные свойства можно проверить непосредственно (рекомендуем это сделать самостоятельно). В дальнейшем (см. § 28) будет показан некоторый символический метод, который, в частности, можно использовать и для установления этих свойств вихря.

В математическом анализе была получена формула Стокса:

$$\oint_L a_x dx + a_y dy + a_z dz =$$

$$= \int\int_S \left(\frac{\partial a_z}{\partial y} - \frac{\partial a_y}{\partial z}\right) dydz + \left(\frac{\partial a_x}{\partial z} - \frac{\partial a_z}{\partial x}\right) dxdz +$$

$$+ \left(\frac{\partial a_y}{\partial x} - \frac{\partial a_x}{\partial y}\right) dxdy.$$

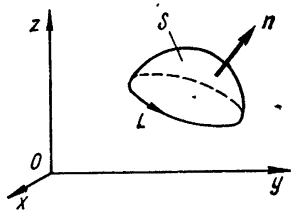

Рис. 74.

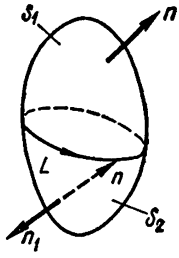

Рис. 75

В этой формуле L — замкнутый контур, S — поверхность, ограниченная этим контуром. Положительное напряжение на контуре L и выбор стороны у поверхности S согласованы (рис. 74). Интеграл в левой части этой формулы равен $\oint_L a_\tau dl$ — циркуляции вектора $\boldsymbol{a}$ вдоль контура L. Интеграл в правой части представляет поток вектора $\operatorname{rot} \boldsymbol{a}$ через поверхность S, ограниченную контуром L (в направлении нормалей, указанных на рис. 74).

Следовательно, формула Стокса запишется так:

$$\oint_L a_\tau dl = \int\int_S \operatorname{rot}_n \boldsymbol{a} dS.$$

Циркуляция вектора $\boldsymbol{a}$ *вдоль замкнутого контура* L *равна потоку вихря этого вектора* $\boldsymbol{a}$ *через поверхность* S, *лежащую в поле вектора* $\boldsymbol{a}$ *и ограниченную контуром* L.

Из теоремы Стокса следует, что поток вихря вектора $\boldsymbol{a}$ не зависит от вида поверхности S, проходящей через контур L. Если через контур L провести две поверхности S_1 и S_2, то

$$\int\int_{S_1} \operatorname{rot}_n \boldsymbol{a} dS = \int\int_{S_2} \operatorname{rot}_n \boldsymbol{a} dS.$$

Поверхности S_1 и S_2 ограничивают некоторый объем V (рис. 75). В предыдущей формуле нормаль к поверхности S_1 — внешняя, нормаль к поверхности S_2 — внутренняя (на рис. 75 показана пунктирной линией). Изменив направление нормали на поверхности S_2 на противоположное (так, чтобы она стала внешней), получим

$$\oint\limits_{S_1+S_2} \operatorname{rot}_n a dS = 0.$$

Следовательно, поток вихря вектора $\boldsymbol{a}$ через замкнутую поверхность равен нулю. Если в некоторой части поля $\boldsymbol{a}$ (во всем поле) $\operatorname{rot} \boldsymbol{a} \equiv 0$, то

$$\oint\limits_L a_\tau dl = \int\int\limits_S 0 \cdot dS = 0,$$

где L — любой контур, лежащий целиком в указанной части поля (во всем поле).

Пример. Постоянный электрический ток I течет вверх по бесконечному длинному проводу, совпадающему с осью Oz (рис. 76). Найти вихрь вектора $\boldsymbol{H}$ напряженности магнитного поля, создаваемого этим током.

Как уже было показано (см. с. 152), в точке $M(x, y, z)$

$$\boldsymbol{H} = \frac{I}{2\pi\rho^2}(-y\boldsymbol{i} + x\boldsymbol{j}),$$

где $\rho^2 = x^2 + y^2$.

Запишем проекции вектора $\boldsymbol{H}$ на оси координат:

$$H_x = -\frac{Iy}{2\pi\rho^2}, \quad H_y = \frac{Ix}{2\pi\rho^2}, \quad H_z = 0.$$

Так как $H_z = 0$, то вектор $\boldsymbol{H}$ лежит в плоскости, перпендикулярной к проводу (к оси Oz). Модуль вектора $\boldsymbol{H}$ будет равен

$$H = \sqrt{H_x^2 + H_y^2 + H_z^2} = \frac{I}{2\pi\rho^2}\sqrt{y^2 + x^2} = \frac{I}{2\pi\rho}.$$

Найдем теперь

$$\operatorname{rot} \boldsymbol{H} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ -\dfrac{Iy}{2\pi\rho^2} & \dfrac{Ix}{2\pi\rho^2} & 0 \end{vmatrix} =$$

$$= \left(\frac{\partial(0)}{\partial y} - \frac{\partial}{\partial z}\left(\frac{Ix}{2\pi\rho^2}\right)\right)\boldsymbol{i} + \left(\frac{\partial}{\partial z}\left(-\frac{Iy}{2\pi\rho^2}\right) - \frac{\partial(0)}{\partial x}\right)\boldsymbol{j} +$$

$$+ \left(\frac{\partial}{\partial x}\left(\frac{Ix}{2\pi\rho^2}\right) - \frac{\partial}{\partial y}\left(-\frac{Iy}{2\pi\rho^2}\right)\right)\boldsymbol{k} = 0 \cdot \boldsymbol{i} + 0 \cdot \boldsymbol{j} +$$

$$+ \frac{I}{2\pi}\left(\frac{\partial}{\partial x}\left(\frac{x}{x^2+y^2}\right) + \frac{\partial}{\partial y}\left(\frac{y}{x^2+y^2}\right)\right)\boldsymbol{k} =$$

$$= \frac{I}{2\pi}\left(\frac{1}{x^2+y^2} - \frac{2x^2}{(x^2+y^2)^2} + \frac{1}{x^2+y^2} + \frac{2y^2}{(x^2+y^2)^2}\right)\boldsymbol{k} =$$

$$= \frac{I}{2\pi(x^2+y^2)^2}(x^2 + y^2 - 2x^2 + x^2 + y^2 - 2y^2)\,\boldsymbol{k} = 0.$$

Итак, rot $\boldsymbol{H} = 0$ во всех точках, не лежащих на проводе. В точках, лежащих на проводе, формулы для H_x, H_y теряют смысл, так как в этих точках $\rho = 0$. Циркуляция вектора $\boldsymbol{H}$ по любой замкнутой кривой, не охватывающей ось Oz, равна нулю. Если же ось Oz проходит внутри контура L, то циркуляция не зависит от вида этого контура. Нетрудно показать, что в этом случае она равна I.

Выше был рассмотрен идеальный случай, когда не учитывались размеры сечения провода с постоянным током I. Если же учесть, что постоянный ток распределяется по сечению провода S равномерно с плотностью $\delta = \frac{I}{S}$, а плотность тока проводимости является вектором, совпадающим по направлению с направлением движения положительных зарядов, то в любой точке внутри провода получим, что

$$\text{rot}\,\boldsymbol{H} = \boldsymbol{\delta}.$$

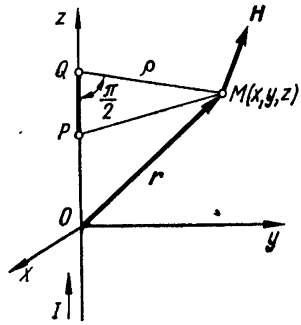

Рис. 76

В теоретической электротехнике доказано, что это уравнение справедливо в любой точке, где имеется отличный от нуля вектор плотности полного тока $\boldsymbol{\delta}$.

Под вектором плотности полного тока понимается геометрическая сумма векторов плотности токов: проводимости $\boldsymbol{\delta}_{пр}$, существующего в твердых и жидких проводящих средах; переноса $\boldsymbol{\delta}_{пер}$, существующего в вакууме и сильно разряженных газах; и смещения в диэлектрике $\boldsymbol{\delta}_{см}$, существующего только при изменении во времени электрического поля, т. е.

$$\boldsymbol{\delta} = \boldsymbol{\delta}_{пр} + \boldsymbol{\delta}_{пер} + \boldsymbol{\delta}_{см}.$$

Тогда

$$\text{rot}\,\boldsymbol{H} = \boldsymbol{\delta}_{пр} + \boldsymbol{\delta}_{пер} + \boldsymbol{\delta}_{см}.$$

Эта формула представляет собой первое уравнение Максвелла, отражающее связь и взаимную обусловленность магнитного и электрического полей, как двух сторон единого электромагнитного поля, являющегося видом материи.

Физический смысл этого уравнения можно выразить так: *магнитное поле может быть возбуждено не только движущимися электрическими зарядами (токами проводимости и переноса), но и изменением электрического поля во времени (током смещения).*

В электромагнитном поле всегда справедливо и второе уравнение Максвелла

$$\text{rot}\,\boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t}, \tag{8}$$

которое также отражает связь и взаимную обусловленность электрического и магнитного полей. Физический смысл уравнения (8) состоит в том, что при изменении во времени магнитного поля (вектора магнитной индукции $\boldsymbol{B}$) в том же пространстве возникает поле электрическое, тесно связанное с магнитным полем и его изменением.

Первое и второе уравнения Максвелла являются основными уравнениями в полной системе уравнений переменного электромагнитного поля.

§ 28. ОПЕРАТОР ГАМИЛЬТОНА И ЕГО ПРИМЕНЕНИЕ

Выше были рассмотрены основные дифференциальные операции (действия) над скалярным полем U и векторным полем

$$\boldsymbol{a} = a_x\boldsymbol{i} + a_y\boldsymbol{j} + a_z\boldsymbol{k}.$$

1. $\operatorname{grad} v = \frac{\partial v}{\partial x} i + \frac{\partial v}{\partial y} j + \frac{\partial v}{\partial z} k.$

2. $\operatorname{div} a = \frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z}.$

3. $\operatorname{rot} a = \begin{vmatrix} i & j & k \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ a_x & a_y & a_z \end{vmatrix} = \left(\frac{\partial a_z}{\partial y} - \frac{\partial a_y}{\partial z}\right) i + \left(\frac{\partial a_x}{\partial z} - \frac{\partial a_z}{\partial x}\right) j +$

$+ \left(\frac{\partial a_y}{\partial x} - \frac{\partial a_x}{\partial y}\right) k.$

Английский математик и механик Гамильтон показал, что эти операции можно записать проще, если применить оператор

$$\nabla = i \cdot \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} + k \cdot \frac{\partial}{\partial z}.$$

Этот оператор по своей природе векторно-дифференциальный: при его применении оператор ∇ сохраняет как свойства вектора, так и свойства оператора дифференцирования. Он называется оператором Гамильтона или оператором «набла». Применяя этот оператор, получим:

1. $\nabla U = i \cdot \frac{\partial U}{\partial x} + j \cdot \frac{\partial U}{\partial y} + k \cdot \frac{\partial U}{\partial z} = \operatorname{grad} U;$

2. $\nabla \cdot a = \frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z} = \operatorname{div} a;$

3. $\nabla \times a = \begin{vmatrix} i & j & k \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ a_x & a_y & a_z \end{vmatrix} = \operatorname{rot} a.$

Так, в частности, в электростатическом поле для градиента электрического потенциала V можно записать:

$$\nabla V = \operatorname{grad} V.$$

Теорема Гаусса запишется в виде:

$$\nabla \cdot E = \frac{\rho}{\varepsilon_a},$$

где ρ — объемная плотность электрического заряда в данной точке поля.

Первое уравнение Максвелла будет иметь вид

$$\nabla \times H = \delta.$$

Принципы непрерывности магнитного поля и электрического поля токов соответственно будут выражены равенствами:

$$\nabla \cdot B = 0 \text{ и } \nabla \cdot \delta = 0.$$

Оператор Гамильтона применяется для записи многих других операций и для вывода различных формул в теории поля. Как дифференциальный оператор он действует только на множитель, который стоит непосредственно за ним. Например,

$$\nabla Uv = (\nabla U)\, v = v \cdot \operatorname{grad} U;$$

$$\nabla (Uv) = \operatorname{grad} (Uv).$$

Если в каком-либо выражении за оператором ∇ нет множителей, то это выражение представляет собой некоторый новый оператор. Так,

$$\boldsymbol{a} \cdot \nabla = a_x \frac{\partial}{\partial x} + a_y \frac{\partial}{\partial y} + a_z \frac{\partial}{\partial z}$$

— это скалярный дифференциальный оператор, который может действовать как на скалярное, так и на векторное поле. Например,

$$(\boldsymbol{a} \cdot \nabla) U = a_x \frac{\partial U}{\partial x} + a_y \frac{\partial U}{\partial y} + a_z \frac{\partial U}{\partial z} = \boldsymbol{a} \cdot \operatorname{grad} U;$$

$$(\boldsymbol{a} \cdot \nabla)\, \boldsymbol{b} = \left(a_x \frac{\partial}{\partial x} + a_y \frac{\partial}{\partial y} + a_z \frac{\partial}{\partial z}\right)(b_x \boldsymbol{i} + b_y \boldsymbol{j} + b_z \boldsymbol{k})$$

— вектор, который получим, если к проекциям вектора $\boldsymbol{b}$ применить операцию

$$a_x \frac{\partial}{\partial x} + a_y \frac{\partial}{\partial y} + a_z \frac{\partial}{\partial z}.$$

Этот вектор называют градиентом векторного поля $\boldsymbol{b}$ в направлении вектора $\boldsymbol{a}$. Например, проекция этого вектора $(\boldsymbol{a} \cdot \nabla)\, \boldsymbol{b}$ на ось Ox будет

$$a_x \frac{\partial b_x}{dx} + a_y \frac{\partial b_x}{\partial y} + a_z \frac{\partial b_x}{\partial z}.$$

Аналогично находятся и проекции на другие оси.

При действиях с оператором Гамильтона надо пользоваться правилами векторной алгебры и правилами дифференцирования. Покажем это на нескольких примерах:

$$\operatorname{grad} (U + v) = \nabla (U + v) = \nabla U + \nabla v = \operatorname{grad} U + \operatorname{grad} v.$$

$$\operatorname{grad} (C \cdot U) = \nabla (C \cdot U) = C \cdot \nabla U = C \cdot \operatorname{grad} U$$

(C — постоянная, ее можно вынести за знак оператора ∇ — по свойству операции дифференцирования);

$$\operatorname{grad} (Uv) = \nabla (Uv) = \nabla Uv + U \nabla v = v \operatorname{grad} U + U \operatorname{grad} v.$$

В данном случае оператор ∇ применяется как оператор дифференцирования к произведению Uv.

$$\operatorname{grad} (\boldsymbol{a} \cdot \boldsymbol{b}) = \nabla (\boldsymbol{a} \cdot \boldsymbol{b}) = \nabla (\boldsymbol{a} \cdot \boldsymbol{b}_c) + \nabla (\boldsymbol{a}_c \cdot \boldsymbol{b}) =$$

$$= (\boldsymbol{b} \cdot \nabla)\, \boldsymbol{a} + \boldsymbol{b} \times (\nabla \times \boldsymbol{a}) + (\boldsymbol{a} \cdot \nabla)\, \boldsymbol{b} + \boldsymbol{a} \times (\nabla \times \boldsymbol{b}) =$$

$$= (\boldsymbol{b} \cdot \operatorname{grad})\, \boldsymbol{a} + \boldsymbol{b} \times \operatorname{rot} \boldsymbol{a} + (\boldsymbol{a} \cdot \operatorname{grad})\, \boldsymbol{b} + \boldsymbol{a} \times \operatorname{rot} \boldsymbol{b}.$$

Формулы $\nabla(a \cdot b_c) = (b \cdot \nabla)\, a + b \times (\nabla \times a)$ и $\nabla(a_c \cdot b) = (a \cdot \nabla)\, b + a \times (\nabla \times b)$ получены на основании выражения двойного векторного произведения:

$$A \times (B \times C) = B(A \cdot C) - C(A \cdot B),$$

из которого следует

$$B(A \cdot C) = C(A \cdot B) + A \times (B \times C).$$

Далее,

$$\operatorname{div}(Ua) = \nabla(Ua) = \nabla U \cdot a + U\nabla \cdot a = \\ = \operatorname{grad} U \cdot a + U \operatorname{div} a$$

и

$$\operatorname{rot}(a \times b) = \nabla \times (a \times b) = \nabla \times (a \times b_c) + \nabla \times (a_c \times b).$$

Индекс «c» показывает, что соответствующая величина с этим индексом при применении оператора ∇ считается постоянной. Для дальнейших преобразований применим формулу

$$A \times (B \times C) = B(A \cdot C) - C(A \cdot B),$$

учитывая что оператор ∇ должен действовать на тот множитель, который считается переменным. Получим

$$\operatorname{rot}(a \times b) = \nabla \times (a \times b_c) + \nabla \times (a_c \times b) = \\ = \{A \times (B \times C) + A_1 \times (B_1 \times C_1)\} = \\ = (b_c \cdot \nabla)\, a - b_c(\nabla \cdot a) + a_c(\nabla \cdot b) - (a_c \cdot \nabla)\, b = \\ = a(\nabla \cdot b) - b(\nabla \cdot a) + (b \cdot \nabla)\, a - (a \cdot \nabla)\, b = \\ = a \operatorname{div} b - b \operatorname{div} a + (b \cdot \nabla)\, a - (a \cdot \nabla)\, b.$$

Примечание. Выражение в фигурных скобках написано для того, чтобы проще было применять формулу разложения соответствующего двойного векторного произведения. В ответе индекс «c» у соответствующих величин снят, так как он играл вспомогательную роль — при применении оператора ∇ как оператора дифференцирования (как понимать выражения $(b \cdot \nabla)\, a$ и $(a \cdot \nabla)\, b$, было указано выше).

Мы рассмотрели некоторые дифференциальные операции первого порядка. После их применения к полю получается новое поле, к которому можно вновь применять эти операции. В результате получаются дифференциальные операции второго порядка. При этом надо внимательно следить за тем, чтобы записанные операции имели смысл. Например, выражение $\operatorname{rot}\operatorname{div} a$ лишено смысла, ибо $\operatorname{div} a$ — величина скалярная, а операция «rot» применяется к векторному полю. Аналогично выражение $\operatorname{grad}\operatorname{rot} a$ также не имеет смысла, так как $\operatorname{rot} a$ — вектор, а операция «grad» применяется к скалярному полю.

Имеется лишь пять дифференциальных операций второго порядка:

$$\operatorname{div}\operatorname{grad} U, \quad \operatorname{rot}\operatorname{grad} U, \quad \operatorname{grad}\operatorname{div} a, \\ \operatorname{div}\operatorname{rot} a, \quad \operatorname{rot}\operatorname{rot} a.$$

Применяя оператор ∇ для их записи, получим

$$\operatorname{div}\operatorname{grad} U = \nabla \cdot (\nabla U) = (\nabla \cdot \nabla) U = \nabla^2 U,$$

где $\nabla^2 = \Delta$ — скалярный оператор «дельта» (оператор Лапласа)

$$\Delta = \nabla \cdot \nabla = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2},$$

$$\operatorname{div}\operatorname{grad} U = \Delta U = \frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2}.$$

Для расчета электростатических полей в области, где отсутствуют электрические заряды, используется уравнение Лапласа

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0,$$

которое может быть записано так:

$$\Delta V = \operatorname{div}\operatorname{grad} V = 0,$$

где V — искомый электрический потенциал.

Для расчета электростатических полей в области, где известен закон распределения объемной плотности заряда ρ в диэлектрике с абсолютной диэлектрической проницаемостью ε_a, применяется уравнение Пуассона

$$\Delta V = -\frac{\rho}{\varepsilon_a}.$$

$$\operatorname{rot}\operatorname{grad} U = \nabla \times (\nabla U) = (\nabla \times \nabla) U = 0$$

(известно, что $\boldsymbol{a} \times \boldsymbol{a} = 0$). Это означает, что поле градиента есть поле безвихревое. В частности, в электротехнике

$$\operatorname{rot}\operatorname{grad} V = 0,$$

$$\operatorname{grad}\operatorname{div} \boldsymbol{a} = \nabla (\nabla \cdot \boldsymbol{a}),$$

$$\operatorname{div}\operatorname{rot} \boldsymbol{a} = \nabla (\nabla \times \boldsymbol{a}) = 0,$$

так как в смешанном произведении векторов имеются два одинаковых вектора. Это значит, что поле вихря — поле соленоидальное (в теории электромагнитного поля $\operatorname{div}\operatorname{rot} H = 0$),

$$\operatorname{rot}\operatorname{rot} \boldsymbol{a} = \nabla \times (\nabla \times \boldsymbol{a}) = \nabla (\nabla \cdot \boldsymbol{a}) - (\nabla \cdot \nabla) \boldsymbol{a} = \operatorname{grad}\operatorname{div} \boldsymbol{a} - \Delta \boldsymbol{a}.$$

Здесь $\Delta \boldsymbol{a}$ — результат применения оператора Лапласа к вектору $\boldsymbol{a}$. Это векторная величина

$$\Delta \boldsymbol{a} = \Delta (a_x \boldsymbol{i} + a_y \boldsymbol{j} + a_z \boldsymbol{k}) = \Delta a_x \boldsymbol{i} + \Delta a_y \boldsymbol{j} + \Delta a_z \boldsymbol{k},$$

где Δa_x, Δa_y, Δa_z — операторы Лапласа, примененные соответственно к a_x, a_y, a_z. Например, в электротехнике

$$\operatorname{rot}(\operatorname{rot} \delta) = \operatorname{grad}(\operatorname{div} \delta) - \nabla^2 \delta.$$

Имеется такое правило применения оператора Гамильтона:

При применении оператора ∇ к произведениям скалярных и векторных полей: Uv, $U\boldsymbol{a}$, $\boldsymbol{a} \cdot \boldsymbol{b}$, $\boldsymbol{a} \times \boldsymbol{b}$ надо применить оператор ∇

к каждому из сомножителей отдельно, считая другой постоянным, и результаты сложить, затем каждое полученное слагаемое преобразовать по правилам векторной алгебры так, чтобы оператор ∇ стоял на предпоследнем месте перед переменным множителем.

§ 29. ПОТЕНЦИАЛЬНЫЕ ПОЛЯ

В теории поля и ее приложениях часто рассматриваются так называемые потенциальные и соленоидальные поля. Рассмотрим здесь первые из них.

Определение. *Векторное поле $\boldsymbol{a}$ называется потенциальным, если оно является градиентом некоторого скалярного поля, т. е.*

$$\boldsymbol{a} = -\operatorname{grad} U. \tag{9}$$

Скалярная функция $U = U(x, y, z)$ называется скалярным потенциалом поля $\boldsymbol{a}$.

Заметим, что знак «—» перед grad U в формуле (9) берется для удобства. В некоторых книгах по теории поля его нет. Это не имеет принципиального значения. Однако, изучая литературу, где применяется понятие потенциального векторного поля, надо уточнять, каким образом вводится понятие потенциального поля. В частности, для электростатического поля имеем

$$\boldsymbol{E} = -\operatorname{grad} V.$$

В соответствии с определением потенциала электрического поля V как работы, совершаемой полем по переносу единичного положительного заряда из данной точки поля в точку, потенциал в которой принят равным нулю, знак «—» перед grad V имеет здесь физический смысл. Он означает, что в направлении вектора напряженности электрического поля $\boldsymbol{E}$ электрический потенциал убывает. Встречаются и другие определения потенциального поля. О связи между различными определениями потенциального поля будет сказано ниже.

Из формулы (9) следует, что для потенциального поля

$$\boldsymbol{a} = a_x\boldsymbol{i} + a_y\boldsymbol{j} + a_z\boldsymbol{k},$$

$$a_x = -\frac{\partial U}{\partial x}, \quad a_y = -\frac{\partial U}{\partial y}, \quad a_z = -\frac{\partial U}{\partial z}.$$

В некоторых книгах потенциальные силовые поля называют консервативными.

Заметим, что если поле $\boldsymbol{a}$ — потенциальное, то его потенциал U не является единственным: функция $U + C$, где C — произвольная постоянная, также является потенциалом этого поля. Последнее следует из того, что

$$\operatorname{grad}(U + C) = \operatorname{grad} U.$$

Учитывая, что в общее выражение потенциала входит произвольная постоянная, можно добиться того, чтобы потенциал поля в некоторой заданной точке был равен нулю. Покажем, что потенциальное поле $\boldsymbol{a}$ является безвихревым:

$$\operatorname{rot} \boldsymbol{a} = 0.$$

Действительно, для потенциального поля (иногда говорят потенциального вектора)

$$\boldsymbol{a} = -\operatorname{grad} U.$$

Тогда

$$\operatorname{rot} \boldsymbol{a} = \operatorname{rot}(-\operatorname{grad} U) = -\operatorname{rot} \operatorname{grad} U = 0,$$

так как rot grad $U = 0$ (см. § 28).

Следовательно,

$$\operatorname{rot} \boldsymbol{a} = 0.$$

Покажем теперь, что в потенциальном поле $\boldsymbol{a}$ циркуляция вектора $\boldsymbol{a}$ по любому замкнутому контуру L, лежащему в этом поле, равна нулю. Для доказательства применим формулу Стокса:

$$\oint_L a_\tau dl = \iint_S \operatorname{rot}_n \boldsymbol{a} dS.$$

Так как поле вектора $\boldsymbol{a}$ — потенциальное, то rot $\boldsymbol{a} = 0$. Поэтому $\operatorname{rot}_n \boldsymbol{a} = 0$ на любой поверхности S, целиком лежащей в этом поле. Следовательно,

$$\oint_L a_\tau dl = 0.$$

Из доказанного следует, что в потенциальном поле циркуляция не зависит от вида кривой, соединяющей две данные точки, а зависит лишь от положения этих точек в поле.

Если, в частности, потенциальное поле — силовое, то для него справедливо следующее: *в потенциальном силовом поле работа силы по любому замкнутому контуру равна нулю.* В этом случае работа силы при переходе из одной точки в другую не зависит от вида кривой, соединяющей эти точки. Так, в электростатическом поле

$$\operatorname{rot} \boldsymbol{E} = 0 \text{ и } \oint_L E_\tau dl = 0.$$

Следует заметить, что в приведенных выше рассуждениях предполагается, что область, где задано векторное поле $\boldsymbol{a}$, является о д н о с в я з н о й. Это означает, что любой замкнутый контур, целиком расположенный в этой области, можно путем непрерывной деформации стянуть в точку (оставаясь в этой области). Справедлива теорема:

Теорема. *Следующие четыре свойства векторного поля $\boldsymbol{a}$, заданного в односвязной области D, эквивалентны:*

1) *Циркуляция поля* $\boldsymbol{a}$ *по любому замкнутому контуру, расположенному в области* D, *равна нулю.*

2) *Циркуляция поля* $\boldsymbol{a}$ *вдоль любой кривой* L *(лежащей в области* D*) с началом в точке* A *и концом в точке* B *зависит только от положения точек* A *и* B *и не зависит от формы кривой.*

3) *Существует функция* $U(x, y, z)$ *такая, что*

$$\boldsymbol{a} = -\operatorname{grad} U.$$

4) *Поле* $\boldsymbol{a}$ *безвихревое, т. е.*

$$\operatorname{rot} \boldsymbol{a} = 0.$$

Предполагается, что проекции $a_x(x, y, z)$, $a_y(x, y, z)$, $a_z(x, y, z)$ вектора $\boldsymbol{a}$ непрерывны вместе со своими первыми частными производными в области D.

Доказательство этой теоремы имеется во многих учебниках и пособиях по векторному анализу.

Эта теорема дает возможность различным образом вводить понятие потенциального поля. Любой из пунктов 1), 2), 3), 4) можно взять в качестве определения потенциального поля. Тогда остальные три пункта будут являться свойствами потенциального поля. В качестве определения потенциального поля $\boldsymbol{a}$ мы приняли то, что записано в пункте 3). Исходя из этого определения и известных теорем анализа, установили свойства 1), 2), 4) потенциального поля. Это определение, очевидно, эквивалентно такому определению потенциального поля, которое часто встречается, например, в теоретической механике.

Поле $\boldsymbol{a}$ *называется* п о т е н ц и а л ь н ы м, *если существует такая функция* $U(x, y, z)$, *что*

$$\frac{\partial U}{\partial x} = -a_x, \quad \frac{\partial U}{\partial y} = -a_y, \quad \frac{\partial U}{\partial z} = -a_z. \tag{10}$$

При этом знаки «—» в написанных выше равенствах иногда и не пишут. Знак минус в формулах (9) и (10) берут потому, что обычно векторные линии направлены в сторону убывания U.

Так, например, электрический ток течет от точек с бóльшим потенциалом к точкам с меньшим потенциалом; теплота перемещается от места, более нагретого, к месту, менее нагретому; поток жидкости направлен туда, где давление меньше, и т. д. Заметим, что в неодносвязной области может быть безвихревое, но не потенциальное поле. Отметим также, что потенциальное поле не может иметь замкнутых векторных линий, ибо вдоль такой линии циркуляция отлична от нуля (проекция a_τ вектора $\boldsymbol{a}$ на направление касательной к этой линии, очевидно, сохраняет знак).

Поверхности уровня потенциала $U(x, y, z)$ поля $\boldsymbol{a}$ называются э к в и п о т е н ц и а л ь н ы м и п о в е р х н о с т я м и. На каждой из них потенциал поля $\boldsymbol{a}$ имеет постоянное значение.

Векторные линии потенциального поля о р т о г о н а л ь н ы эквипотенциальным поверхностям. Это следует из того, что grad U

направлен по нормали к поверхности уровня поля U, а так как $\boldsymbol{a} = -\text{grad}\, U$, то градиент U направлен по касательной к векторным линиям поля $\boldsymbol{a}$.

Рассмотрим теперь вопрос о нахождении потенциала поля.

Пусть поле $\boldsymbol{a}$ — потенциальное. Требуется найти его потенциал, равный нулю в точке M_0.

Возьмем произвольную точку $M(x, y, z)$ этого поля. Вычислим циркуляцию вектора $\boldsymbol{a}$ по произвольной кривой MM_0:

$$\int_{MM_0} a_\tau dl = \int_{MM_0} a_x dx + a_y dy + a_z dz =$$

$$= -\int_{MM_0} \left(\frac{\partial U}{\partial x} dx + \frac{\partial U}{\partial y} dy + \frac{\partial U}{\partial z} dz \right) = -\int_{MM_0} dU =$$

$$= \int_{M_0 M} dU = U \Big|_{M_0}^{M} = U(M) - U(M_0).$$

Отсюда $U(M) = \int_{MM_0} a_x dx + a_y dy + a_z dz + U(M_0)$, но по условию $U(M_0) = 0$, тогда

$$U(M) = \int_{MM_0} a_x dx + a_y dy + a_z dz. \qquad (11)$$

Из этой формулы следует, что *скалярный потенциал поля $\boldsymbol{a}$ в точке $M(x, y, z)$ равен циркуляции вектора $\boldsymbol{a}$ вдоль произвольной кривой, соединяющей точку M с точкой M_0, в которой потенциал поля принят равным нулю.*

Если потенциальное поле $\boldsymbol{a}$ — силовое, то потенциал этого поля в точке $M(x, y, z)$ численно равен работе силы вдоль произвольной кривой, соединяющей точку M с точкой нулевого потенциала M_0.

В некоторых простейших случаях вопрос о том, будет ли данное поле потенциальным, решается путем непосредственного нахождения скалярной функции $U(x, y, z)$, удовлетворяющей условиям (9) или (10). Например, поле $\boldsymbol{a} = \boldsymbol{r} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$, очевидно, потенциальное, его потенциалом, равным нулю в начале координат, будет функция

$$U(x, y, z) = -\frac{x^2 + y^2 + z^2}{2}.$$

(Рекомендуем самостоятельно проверить выполнение равенств (10)).

Однако в ряде других случаев таким способом решить вопрос о том, будет ли данное поле потенциальным, не просто. Тогда поступают иначе. Находят вихрь данного поля. Если он равен нулю во всех точках некоторой односвязной области (или во всем пространстве), то поле в этой области (во всем пространстве) будет потенциальным. После этого, выбрав точку, где потенциал должен быть равен нулю, находят потенциал в произвольной точке поля по формуле (11).

Примеры

1. Убедиться, что поле силы тяжести потенциальное, найти его потенциал, равный нулю в начале координат.

Р е ш е н и е. Поле силы тяжести задается вектором

$$\boldsymbol{P} = -mg\boldsymbol{k},$$

где m — масса материальной точки, помещенной в этом поле; g — ускорение силы тяжести; $\boldsymbol{k}$ — орт оси Oz.

В данном случае $P_x = 0$, $P_y = 0$, $P_z = -mg$.

Найдем

$$\operatorname{rot} \boldsymbol{P} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ 0 & 0 & -mg \end{vmatrix} = 0.$$

Так как $\operatorname{rot} \boldsymbol{P} = 0$ во всех точках $M(x, y, z)$, то поле силы тяжести — потенциальное во всем пространстве. Найдем его потенциал, равный нулю в начале координат:

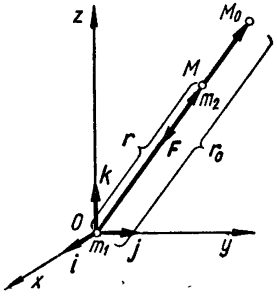

Рис. 77

$$U(x, y, z) = \int_{MO} P_x dx + P_y dy + P_z dz =$$

$$= -\int_{MO} mg dz = -mgz\Big|_z^0 = mgz = Pz.$$

2. Убедиться, что гравитационное поле — потенциальное, и найти его потенциал, равный нулю на бесконечности.

Р е ш е н и е. Гравитационное поле — поле силы взаимодействия между двумя материальными точками. Если массы этих точек m_1 и m_2, расстояние между ними r, то модуль силы взаимодействия F находится по формуле

$$F = \frac{\gamma m_1 m_2}{r^2},$$

где γ — постоянная величина (гравитационная постоянная).

Предположим, что точка m_1 находится в начале координат и F — сила, с которой точка m_1 действует на точку m_2 (рис. 77). Тогда

$$\boldsymbol{F} = -\frac{\gamma m_1 m_2}{r^2} \cdot \frac{\boldsymbol{r}}{r};$$

$$F_x = -\frac{\gamma m_1 m_2 x}{r^3}, \quad F_y = -\frac{\gamma m_1 m_2 y}{r^3}, \quad F_z = -\frac{\gamma m_1 m_2 z}{r^3},$$

где

$$r = \sqrt{x^2 + y^2 + z^2}, \quad \boldsymbol{r} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}.$$

Найдем

$$\operatorname{rot} \boldsymbol{F} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ F_x & F_y & F_z \end{vmatrix} = \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right)\boldsymbol{i} +$$

$$+\left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}\right)\boldsymbol{j} + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right)\boldsymbol{k}.$$

Вычислим

$$\frac{\partial F_z}{\partial y} = -\gamma m_1 m_2 z \cdot \frac{\partial (r^{-3})}{\partial y} = \frac{3\gamma m_1 m_2 z}{r^4} \cdot \frac{\partial r}{\partial y} =$$

$$= \frac{3\gamma m_1 m_2 z}{r^4} \cdot \frac{y}{r} = \frac{3\gamma m_1 m_2 zy}{r^5}$$

и

$$\frac{\partial F_y}{\partial z} = \frac{3\gamma m_1 m_2 yz}{r^5}.$$

Тогда

$$\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} = 0.$$

Аналогично получим, что

$$\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} = 0, \quad \frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} = 0.$$

Следовательно,

$$\operatorname{rot} F = 0.$$

Поле силы F потенциально. Оно определено во всем пространстве, за исключением одной точки — начала координат, где помещена масса m_1 (в этой точке $r = 0$).

Найдем потенциал гравитационного поля, равный нулю на бесконечности (это часто требуется, например, в теоретической механике). На луче, проходящем через точку M из начала координат, возьмем точку M_0 с радиусом-вектором r_0 (рис. 77). По формуле (11) получим

$$U(M) = \int_{MM_0} F_x dx + F_y dy + F_z dz = -\gamma m_1 m_2 \int_{MM_0} \frac{r dr}{r^3} =$$

$$= -\gamma m_1 m_2 \int_{MM_0} \frac{dr}{r^2} = \gamma m_1 m_2 \cdot \frac{1}{r}\Big|_r^{r_0} = \gamma m_1 m_2 \left(\frac{1}{r_0} - \frac{1}{r}\right).$$

Предположим теперь, что точка M_0 по лучу неограниченно удаляется от начала координат ($r_0 \to \infty$). Тогда

$$U(M) = -\frac{\gamma m_1 m_2}{r}. \tag{12}$$

Это и будет потенциал гравитационного поля в точке M, который равен нулю на бесконечности.

Совершенно аналогично может быть решена задача для электростатического поля двух точечных электрических зарядов Q_1 и $Q_2 = +1$ Кл, помещенных соответственно в точках 0 и M в однородном и изотропном диэлектрике с абсолютной диэлектрической проницаемостью ε_a.

Модуль силы взаимодействия E между этими зарядами определяется по закону Кулона:

$$E = \frac{Q_1 \cdot 1}{4\pi\varepsilon_a \cdot r^2},$$

а вектор силы запишется в виде

$$\boldsymbol{E} = -\frac{Q_1}{4\pi\varepsilon_a r^2}\cdot\frac{\boldsymbol{r}}{r}.$$

В результате получим

$$\operatorname{rot}\boldsymbol{E} = 0, \quad V(M) = \frac{Q_1}{4\pi\varepsilon_a}\left(\frac{1}{r_0} - \frac{1}{r}\right)$$

и при $r_0 \to \infty$

$$V(M) = -\frac{Q_1}{4\pi\varepsilon_a r}.$$

Последнее равенство выражает потенциал электрического поля отрицательного точечного заряда в текущей точке M.

Если бы заряд Q_1 был положительным, то поле совершало бы положительную работу по «выталкиванию» заряда $Q_2 = +1$ Кл в бесконечность, где электрический потенциал принят равным нулю. Тогда мы получили бы и значение потенциала в точке M также положительным.

Заметим, что в электрических цепях за точку нулевого потенциала можно принять любую точку цепи (а не точку, лежащую в бесконечности). Обычно принимают электрический потенциал равным нулю в точке заземления (в точке цепи, соединенной с землей). Тогда потенциалы остальных точек цепи определяют относительно этой точки заземления.

Потенциалы вида (12) встречаются довольно часто, поэтому целесообразно изучить их несколько подробней. С этой целью рассмотрим следующий пример:

Пример. Пусть дано потенциальное поле, имеющее потенциал вида

$$U(x, y, z) = \frac{e}{r}, \quad \text{где } r = \sqrt{x^2 + y^2 + z^2}.$$

Такое поле называется полем источника, расположенного в начале координат. Это поле определено всюду, исключая начало координат, где U обращается в бесконечность.

По определению потенциала вектор $\boldsymbol{a}$, задающий поле, запишем в виде

$$\boldsymbol{a} = -\operatorname{grad} U = -\operatorname{grad}\frac{e}{r} = -e\operatorname{grad}\left(\frac{1}{r}\right) =$$

$$= -e\left(\frac{\partial}{\partial x}\left(\frac{1}{r}\right)\boldsymbol{i} + \frac{\partial}{\partial y}\left(\frac{1}{r}\right)\boldsymbol{j} + \frac{\partial}{\partial z}\left(\frac{1}{r}\right)\boldsymbol{k}\right) =$$

$$= e\frac{x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}}{r^3} = \frac{e\boldsymbol{r}}{r^3}.$$

Найдем расходимость поля $\boldsymbol{a}$:

$$\operatorname{div}\boldsymbol{a} = \operatorname{div}\left(-\operatorname{grad}\frac{e}{r}\right) = -e\operatorname{div}\operatorname{grad}\left(\frac{1}{r}\right) =$$

$$= -e\Delta\left(\frac{1}{r}\right) = -e\left(\frac{\partial^2}{\partial x^2}\left(\frac{1}{r}\right) + \frac{\partial^2}{\partial y^2}\left(\frac{1}{r}\right) + \frac{\partial^2}{\partial z^2}\left(\frac{1}{r}\right)\right).$$

$$\frac{\partial^2}{\partial x^2}\left(\frac{1}{r}\right) = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}\left(\frac{1}{r}\right)\right) = \frac{\partial}{\partial x}\left(-\frac{1}{r^2}\cdot\frac{\partial}{\partial x}(r)\right) =$$

$$= \frac{\partial}{\partial x}\left(-\frac{1}{r^2}\cdot\frac{x}{r}\right) = -\frac{\partial}{\partial x}\left(\frac{x}{r^3}\right) = -\left(\frac{1}{r^3} + x\frac{\partial}{\partial x}(r^{-3})\right) =$$

$$= -\left(\frac{1}{r^3} - 3xr^{-4}\frac{\partial}{\partial x}(r)\right) = -\left(\frac{1}{r^3} - 3xr^{-4}\cdot\frac{x}{r}\right) =$$

$$= \frac{3x^2}{r^5} - \frac{1}{r^3} = \frac{3x^2 - r^2}{r^5} = \frac{3x^2 - (x^2+y^2+z^2)}{r^5} = \frac{2x^2 - y^2 - z^2}{r^5}.$$

Аналогично

$$\frac{\partial^2}{\partial y^2}\left(\frac{1}{r}\right) = \frac{2y^2 - z^2 - x^2}{r^5},$$

$$\frac{\partial^2}{\partial z^2}\left(\frac{1}{r}\right) = \frac{2z^2 - x^2 - y^2}{r^5}.$$

Подставив эти производные в выражение для div $\boldsymbol{a}$, получим

$$\operatorname{div}\boldsymbol{a} = 0.$$

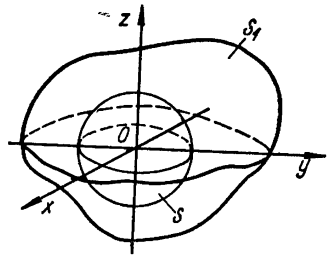

Рис. 78

По теореме Остроградского найдем, что *поток поля через любую замкнутую поверхность, не содержащую начало координат, равен нулю.* Если поверхность охватывает начало координат, то теорему Остроградского применить нельзя, так как в этой точке поле и его расходимость теряют смысл.

Рассмотрим сферу S радиуса R с центром в начале координат (рис. 78). Вычислим поток поля через эту сферу в направлении внешней нормали к ней:

$$\oint\limits_{S^+} a_n dS.$$

В данном случае a_n постоянна на сфере. Действительно,

$$a_n = \operatorname{Пр}_n\boldsymbol{a} = \frac{\boldsymbol{a}\cdot\boldsymbol{r}}{r}\Big|_{r=R} = \frac{er^2}{r^4}\Big|_{r=R} = \frac{e}{R^2}.$$

Итак,

$$\oint\limits_{S^+} a_n dS = \frac{e}{R^2}\cdot 4\pi R^2 = 4\pi e.$$

Возьмем произвольную замкнутую поверхность S_1, охватывающую начало координат и сферу S с центром в начале координат.

Пусть V — тело, ограниченное поверхностями S и S_1.

В теле V имеем div $\boldsymbol{a} = 0$, следовательно, поток поля $\boldsymbol{a}$ через поверхность этого тела равен 0:

$$\oint\limits_{S_1^+, S^-} a_n dS = 0,$$

где S_1^+ — внешняя сторона, S^- — внутренняя сторона поверхностей S_1 и S соответственно.

Тогда

$$\iint\limits_{S_1^+} a_n dS + \iint\limits_{S^-} a_n dS = 0,$$

$$\iint\limits_{S_1^+} a_n dS = -\iint\limits_{S^-} a_n dS = \oint\limits_{S^+} a_n dS = 4\pi e.$$

Таким образом, *поток поля источника через замкнутую поверхность равен нулю, если поверхность не охватывает источник, и равен* $4\pi e$, *если поверхность охватывает источник.* Для электростатического поля, созданного положительным зарядом Q в диэлектрике с ε_a, величина e равна $\frac{Q}{4\pi\varepsilon_a}$. Поэтому

$$\oint\limits_{S_1^+} E_n dS = \frac{Q}{\varepsilon_a},$$

что выражает упоминавшуюся ранее теорему Гаусса.

Если поле $\boldsymbol{a}$ есть поле скоростей в установившемся потоке жидкости, то поток поля через поверхность равен количеству жидкости, протекающей через эту поверхность в единицу времени. Величина $4\pi e$ в этом случае представляет собой количество жидкости, даваемое источником в единицу времени. Величина $4\pi e$ называется мощностью (интенсивностью) источника. Если источник расположен не в начале координат, а в точке $M_1(x_1, y_1, z_1)$, то в этом случае

$$U(x, y, z) = -\frac{e}{|r(M) - r(M_1)|} = -\frac{e}{\sqrt{(x-x_1)^2 + (y-y_1)^2 + (z-z_1)^2}},$$

$$\boldsymbol{a} = -\frac{e\,(r(M) - r(M_1))}{|r(M) - r(M_1)|^3},$$

где $\boldsymbol{r}(M)$, $\boldsymbol{r}(M_1)$ — радиусы-векторы точек M и M_1.

§ 30. СОЛЕНОИДАЛЬНЫЕ ПОЛЯ

Определение. *Векторное поле* $\boldsymbol{a}$ *называется соленоидальным, если его расходимость тождественно равна нулю, т. е.*

$$\operatorname{div} \boldsymbol{a} \equiv 0. \tag{13}$$

Соленоидальное поле — поле без источников и стоков.

Для соленоидальных полей справедлива теорема, аналогичная приведенной выше теореме о потенциальных полях.

Теорема. *Следующие четыре свойства векторного поля эквивалентны:*

1) *Поток поля через любую замкнутую поверхность равен нулю.*

2) *Поток поля через поверхность* S, *ограниченную контуром* L *и согласованно с ним ориентированную, зависит только от выбора контура* L *и не зависит от конкретного выбора поверхности* S.

3) *Существует такое поле* $\boldsymbol{A}$, *что*

$$\boldsymbol{a} = \operatorname{rot} \boldsymbol{A}.$$

(Поле $\boldsymbol{A}$ *называется векторным потенциалом поля* $\boldsymbol{a}$*).*

4) *Расходимость поля* $\boldsymbol{a}$ *равна нулю.*

Проведем доказательство теоремы по этапам.

1. Докажем сначала, что из 1) следует 4).

По условию 1) имеем $\oint\limits_S a_n dS = 0$, где S — произвольная замкнутая поверхность. По теореме Остроградского

$$\oint\limits_S a_n dS = \iiint\limits_V \operatorname{div} \boldsymbol{a}\, dv.$$

Тогда

$$\iiint_V \operatorname{div} a dv = 0,$$

где V — объем, ограниченный поверхностью S. Итак, интеграл от расходимости поля по произвольному объему равен нулю, откуда следует, что

$$\operatorname{div} a = 0.$$

2. Докажем, что из 4) следует 3).

Это значит, что нам нужно доказать существование векторного потенциала поля a, т. е. найти поле A такое, что

$$a = \operatorname{rot} A$$

или

$$\left.\begin{aligned} a_x &= \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}, \\ a_y &= \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}, \\ a_z &= \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}, \end{aligned}\right\} \tag{14}$$

при условии, что $\operatorname{div} a = 0$. Будем искать векторный потенциал A, предполагая, что $A_z = 0$ тождественно, $A_y = 0$ на плоскости xOy, т. е. $A_y(x, y, 0) = 0$.

При этом система уравнений (14) примет такой вид

$$\left.\begin{aligned} a_x &= -\frac{\partial A_y}{\partial z}, \\ a_y &= -\frac{\partial A_x}{\partial z}, \\ a_z &= \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}. \end{aligned}\right\} \tag{15}$$

Интегрируя два первых уравнения из (15) по z в пределах от 0 до z, получим

$$\int_0^z a_x(x, y, z)\, dz = -\int_0^z \frac{\partial A_y}{\partial z}\, dz = A_y(x, y, 0) - A_y(x, y, z) =$$

$$= -A_y(x, y, z),$$

$$\int_0^z a_y(x, y, z)\, dz = A_x(x, y, z) - A_x(x, y, 0).$$

Отсюда найдем $A_x(x, y, z)$ и $A_y(x, y, z)$ и подставим их в третье уравнение из (15):

$$a_z = \frac{\partial}{\partial x}\left(-\int_0^z a_x(x, y, z)\,dz\right) - \frac{\partial}{\partial y}\left(\int_0^z a_y(x, y, z)\,dz + A_x(x, y, 0)\right) =$$

$$= -\int_0^z \left(\frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y}\right) dz - \frac{\partial A_x(x, y, 0)}{\partial y}.$$

По условию 4)

$$\operatorname{div} \boldsymbol{a} = \frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z} = 0.$$

Тогда

$$\frac{\partial a_z}{\partial z} = -\left(\frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y}\right).$$

Следовательно,

$$a_z = \int_0^z \frac{\partial a_z}{\partial z}\,dz - \frac{\partial A_x(x, y, 0)}{\partial y} = a_z(x, y, z) - a_z(x, y, 0) -$$

$$- \frac{\partial A_x(x, y, 0)}{\partial y}.$$

Чтобы удовлетворить этому равенству, а значит, и третьему уравнению из (15), достаточно взять

$$A_x(x, y, 0) = -\int_0^y a_z(x, y, z)\,dy$$

(проверьте это, найдя $\frac{\partial A_x}{\partial y}$ и подставив в выражение для a_z).

Таким образом, мы нашли функции:

$$A_x(x, y, z) = \int_0^z a_y(x, y, z)\,dz - \int_0^y a_z(x, y, z)\,dy,$$

$$A_y(x, y, z) = -\int_0^z a_x(x, y, z)\,dz,$$

$$A_z(x, y, z) = 0$$

такие, что векторное поле

$$\boldsymbol{A} = A_x\boldsymbol{i} + A_y\boldsymbol{j} + A_z\boldsymbol{k} = A_x\boldsymbol{i} + A_y\boldsymbol{j}$$

является векторным потенциалом поля $\boldsymbol{a}$.

3. Докажем, что из 3) следует 2).

Пусть L — некоторый контур, S — ограниченная им поверхность. По условию 3) поле $\boldsymbol{a}$ имеет векторный потенциал $\boldsymbol{A}$:

$$\boldsymbol{a} = \operatorname{rot} \boldsymbol{A}.$$

По теореме Стокса

$$\iint\limits_S a_n dS = \iint\limits_S (\operatorname{rot} A)_n \, dS = \oint\limits_L A_\tau dl.$$

Правая часть $\oint\limits_L A_\tau dl$ зависит лишь от выбора контура L, но не зависит от выбора поверхности S, ограниченной этим контуром.

Тогда и левая часть последнего равенства $\iint\limits_S a_n dS$ тоже не зависит от выбора S, что и требовалось доказать.

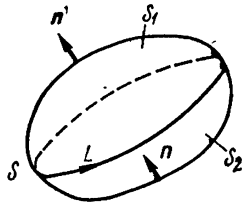

Рис. 79

4. Докажем, что из 2) следует 1).

Возьмем произвольную замкнутую поверхность S. Пусть L — замкнутый контур на S. Контур L разбивает поверхность S на части S_1 и S_2. Выбор сторон у поверхностей S_1 и S_2 согласуем с выбором положительного направления на контуре L (рис. 79).

Из условия 2) следует, что потоки поля $\boldsymbol{a}$ через поверхность S_1 и через поверхность S_2 равны:

$$\iint\limits_{S_1^+} a_n dS = \iint\limits_{S_2^-} a_n dS = -\iint\limits_{S_2^+} a_n dS.$$

Отсюда

$$\iint\limits_{S_1^+} a_n dS + \iint\limits_{S_2^+} a_n dS = 0,$$

$$\iint\limits_{S^+} a_n dS = 0.$$

Итак, поток поля $\boldsymbol{a}$ через произвольную замкнутую поверхность S равен нулю.

Таким образом, теорема полностью доказана.

Заметим, что эта теорема справедлива при некоторых ограничениях на область D, в которой задано поле a. В частности, теорема будет справедлива, если поле $\boldsymbol{a}$ задано во всем пространстве.

Рассмотрим более подробно понятие векторного потенциала.

Отметим, что так же, как и скалярный потенциал потенциального поля, векторный потенциал соленоидального поля не определяется однозначно данным соленоидальным полем. Это следует из того, что поле A и поле $A + \operatorname{grad} \varphi$, где φ — произвольная функция, имеют один и тот же вихрь:

$$\operatorname{rot}(A + \operatorname{grad} \varphi) = \operatorname{rot} A + \operatorname{rotgrad} \varphi = \operatorname{rot} A, \text{ ибо } \operatorname{rotgrad} \varphi \equiv 0.$$

Пользуясь свободой в выборе векторного потенциала A соленоидального поля $\boldsymbol{a}$, можно добиться того, чтобы $\operatorname{div} A = 0$. Это можно сделать так:

Пусть A_1 — векторный потенциал поля a такой, что $\operatorname{div} A_1 \neq 0$. Возьмем в качестве векторного потенциала вектор $A = A_1 + \operatorname{grad} \varphi$, где φ — пока неизвестная скалярная функция. Выберем φ так, чтобы $\operatorname{div} A$ была равна нулю. Тогда

$$\operatorname{div} A = \operatorname{div} A_1 + \operatorname{div} \operatorname{grad} \varphi = \operatorname{div} A_1 + \Delta\varphi = 0.$$

Для нахождения функции φ мы получим уравнение Пуассона

$$\Delta\varphi = -\operatorname{div} A_1.$$

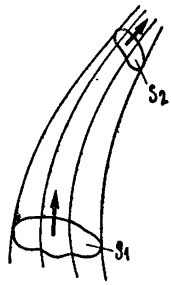

Рис. 80

В качестве решения уравнения Пуассона можно взять функцию

$$\varphi(x, y, z) = \frac{1}{4\pi} \iiint_V \frac{\operatorname{div} A_1(x_1, y_1, z_1)\, dx_1 dy_1 dz_1}{\sqrt{(x-x_1)^2 + (y-y_1)^2 + (z-z_1)^2}},$$

которая является объемным скалярным потенциалом. В электростатике это потенциал поля заряда, распределенного по объему с плотностью $\frac{1}{\varepsilon_a} \operatorname{div} A_1$. В магнитном поле, где всюду $\operatorname{div} B = 0$, принимают $B = \operatorname{rot} A$ и $\operatorname{div} A = 0$. В результате получают выражение векторного потенциала магнитного поля в виде

$$A = \frac{\mu_a}{4\pi} \iiint_V \frac{\delta\, dv}{r},$$

где μ — абсолютная магнитная проницаемость однородной среды, δ — вектор плотности полного тока в текущей точке.

Теперь возьмем векторное поле a.

Векторные линии поля $\operatorname{rot} a$ называются вихревыми линиями поля a. У потенциального поля вихревые линии отсутствуют (потенциальное поле безвихревое!).

Поверхность, образованная векторными линиями поля a, называется векторной трубкой поля a. Очевидно, поток поля через поверхность векторной трубки (или ее часть) равен нулю, так как на поверхности трубки поле направлено вдоль этой поверхности.

В случае соленоидального поля поток поля через любое сечение векторной трубки одинаков. Покажем это.

Пусть S_1 и S_2 — два сечения векторной трубки (рис. 80).

Рассмотрим замкнутую поверхность, образованную сечениями S_1 и S_2 и частью поверхности трубки. Поток поля через всю эту поверхность равен нулю (поле — соленоидальное!). Так как поток через поверхность трубки тоже равен нулю, то поток через поверхность S_1 и поток через поверхность S_2 (в указанных направлениях) одинаковы. Отмеченное свойство имеет место потому, что соленоидальное поле является полем без источников (истоков) и стоков.

§ 31. ОСНОВНЫЕ ОПЕРАЦИИ ВЕКТОРНОГО АНАЛИЗА В КРИВОЛИНЕЙНЫХ КООРДИНАТАХ

Основные понятия теории поля — градиент, дивергенция и вихрь имеют вполне определенный физический смысл, и для введения этих понятий нет необходимости брать ту или иную систему координат. Известно, например, что градиент скалярного поля есть скорость изменения функции, задающей это поле, по тому направлению, по которому эта скорость достигает своего максимума. Дивергенция представляет собой плотность источников или стоков поля. Вихрь является циркуляцией поля, приходящейся на единицу площади поверхности, когда эта поверхность ориентирована так, чтобы циркуляция была наибольшей.

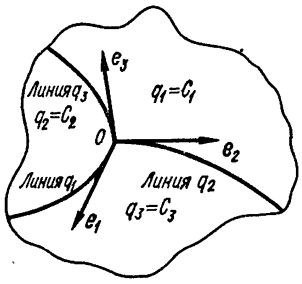

Рис. 81

При конкретном вычислении градиента, расходимости и вихря всегда приходится использовать ту или иную координатную систему. Можно, конечно, всюду пользоваться прямоугольными декартовыми координатами. Однако во многих случаях это связано с большими техническими трудностями. Аналогичное положение, как известно, встречается, например, при вычислении кратных интегралов. И в данном случае часто целесообразно пользоваться криволинейными координатами (например, цилиндрическими или сферическими).

Поэтому рассмотрим здесь вопрос о вычислении градиента, дивергенции и вихря в криволинейных координатах.

Пусть x, y, z — прямоугольные декартовы координаты точки M, q_1, q_2, q_3 — ее криволинейные координаты. При этом предполагается, что каждой тройке q_1, q_2, q_3 соответствует вполне определенная точка пространства и, наоборот, каждой точке пространства соответствует определенная тройка чисел q_1, q_2, q_3. Очевидно,

$$q_1 = q_1(x, y, z), \quad q_2 = q_2(x, y, z), \quad q_3 = q_3(x, y, z),$$

а

$$x = x(q_1, q_2, q_3), \quad y = y(q_1, q_2, q_3), \quad z = z(q_1, q_2, q_3).$$

Уравнения $q_1 = c_1$, $q_2 = c_2$, $q_3 = c_3$, где c_1, c_2, c_3 — постоянные, определяют координатные поверхности. На каждой из них одна из координат q_1, q_2, q_3 остается постоянной.

Предположим, что через каждую точку пространства проходит лишь одна координатная поверхность из каждого семейства. Координатные поверхности попарно пересекаются по линиям, которые называются координатными линиями (рис. 81). Вдоль каждой координатной линии одна из координат q_1, q_2, q_3 меняется, а две другие остаются постоянными. Будем считать, что в каждой точке координатные линии ортогональны друг другу. Пусть e_1, e_2, e_3 — орты касательных к координатным линиям в точке

0, а dl_1, dl_2, dl_3 — элементы длины дуг соответствующих координатных линий.

Известно, что

$$dl = |d\boldsymbol{r}|,$$

где

$$\boldsymbol{r} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}.$$

Вдоль координатной линии q_1 всегда $q_2 = \text{const}$, поэтому x, y, z в точках этой линии зависят лишь от q_1. Тогда

$$d\boldsymbol{r} = \frac{\partial x}{\partial q_1} dq_1 \cdot \boldsymbol{i} + \frac{\partial y}{\partial q_1} dq_1 \cdot \boldsymbol{j} + \frac{\partial z}{\partial q_1} dq_1 \cdot \boldsymbol{k},$$

$$|d\boldsymbol{r}| = \sqrt{\left(\frac{\partial x}{\partial q_1}\right)^2 + \left(\frac{\partial y}{\partial q_1}\right)^2 + \left(\frac{\partial z}{\partial q_1}\right)^2}\, dq_1.$$

Следовательно,

$$dl_1 = H_1 dq_1,$$

где

$$H_1 = \sqrt{\left(\frac{\partial x}{\partial q_1}\right)^2 + \left(\frac{\partial y}{\partial q_1}\right)^2 + \left(\frac{\partial z}{\partial q_1}\right)^2}.$$

Аналогично имеем

$$dl_2 = H_2 dq_2, \quad dl_3 = H_3 dq_3,$$

где

$$H_2 = \sqrt{\left(\frac{\partial x}{\partial q_2}\right)^2 + \left(\frac{\partial y}{\partial q_2}\right)^2 + \left(\frac{\partial z}{\partial q_2}\right)^2},$$

$$H_3 = \sqrt{\left(\frac{\partial x}{\partial q_3}\right)^2 + \left(\frac{\partial y}{\partial q_3}\right)^2 + \left(\frac{\partial z}{\partial q_3}\right)^2}.$$

Элементы площадей координатных поверхностей находятся по формулам:

$$dS_1 = H_2 H_3 dq_2 dq_3,$$

$$dS_2 = H_3 H_1 dq_3 dq_1, \quad dS_3 = H_1 H_2 dq_1 dq_2,$$

H_1, H_2, H_3 называются коэффициентами Ламе.

Найдем выражение градиента функции U в криволинейных координатах q_1, q_2, q_3.

Известно, что

$$\text{Пр}_l \operatorname{grad} U = \frac{\partial U}{\partial l}.$$

Поэтому

$$\text{Пр}_{l_1} \operatorname{grad} U = \frac{\partial U}{\partial l_1}, \quad \text{Пр}_{l_2} \operatorname{grad} U = \frac{\partial U}{\partial l_2},$$

$$\text{Пр}_{l_3} \operatorname{grad} U = \frac{\partial U}{\partial l_3},$$

но

$$\frac{\partial U}{\partial l_1} = \frac{dU}{dl_1} = \frac{dU}{H_1 dq_1} = \frac{1}{H_1} \cdot \frac{\partial U}{\partial q_1}.$$

(Вместо $\frac{\partial U}{dl_1}$ пишем $\frac{\partial U}{\partial q_1}$, потому что функция U в общем случае зависит от q_1, q_2, q_3; вдоль линии q_1 она зависит лишь от q_1.)

Аналогично

$$\frac{\partial U}{\partial l_2} = \frac{1}{H_2} \cdot \frac{\partial U}{\partial q_2}, \quad \frac{\partial U}{\partial l_3} = \frac{1}{H_3} \cdot \frac{\partial U}{\partial q_3}.$$

Следовательно,

$$\operatorname{grad} U = \frac{1}{H_1} \cdot \frac{\partial U}{\partial q_1} \cdot e_1 + \frac{1}{H_2} \cdot \frac{\partial U}{\partial y_2} \cdot e_2 + \frac{1}{H_3} \cdot \frac{\partial U}{\partial q_3} \cdot e_3. \quad (16)$$

Можно показать, что

$$\operatorname{div} a = \frac{1}{H_1 H_2 H_3} \left(\frac{\partial}{\partial q_1} (H_2 H_3 a_{e_1}) + \frac{\partial}{\partial q_2} (H_3 H_1 a_{e_2}) + \right.$$
$$\left. + \frac{\partial}{\partial q_3} (H_1 H_2 a_{e_3}) \right), \quad (17)$$

где $a_{e_i} = \operatorname{Пр}_{e_i} a$ $(i = 1, 2, 3)$.

Если учесть, что оператор Лапласа $\Delta U = \operatorname{div} \operatorname{grad} U$, то, применяя формулы (16) и (17), получим

$$\Delta U = \frac{1}{H_1 H_2 H_3} \left(\frac{\partial}{\partial q_1} \left(\frac{H_2 H_3}{H_1} \frac{\partial U}{\partial q_1} \right) + \frac{\partial}{\partial q_2} \left(\frac{H_3 H_1}{H_2} \frac{\partial U}{\partial q_2} \right) + \right.$$
$$\left. + \frac{\partial}{\partial q_3} \left(\frac{H_1 H_2}{H_3} \frac{\partial U}{\partial q_3} \right) \right). \quad (18)$$

Формула для нахождения вихря вектора a в криволинейных координатах имеет вид

$$\operatorname{rot} a = \frac{1}{H_2 H_3} \left(\frac{\partial}{\partial q_2} (H_3 a_{e_3}) - \frac{\partial}{\partial q_3} (H_2 a_{e_2}) \right) e_1 +$$
$$+ \frac{1}{H_3 H_1} \left(\frac{\partial}{\partial q_3} (H_1 a_{e_1}) - \frac{\partial}{\partial q_1} (H_3 a_{e_3}) \right) e_2 +$$
$$+ \frac{1}{H_1 H_2} \left(\frac{\partial}{\partial q_1} (H_2 a_{e_2}) - \frac{d}{\partial q_2} (H_1 a_{e_1}) \right) e_3. \quad (19)$$

Применим эти формулы для записи основных операций векторного анализа в цилиндрических и сферических координатах.

Цилиндрические координаты. Положение точки M в этой системе координат определяется координатами r, φ, z (рис. 82). При этом $0 \leqslant r < +\infty$, $0 \leqslant \varphi < 2\pi$, $-\infty < z < +\infty$.

Координатные поверхности:

$r = \text{const}$ — цилиндры с осью Oz;

$\varphi = \text{const}$ — полуплоскости, ограниченные осью Oz;

$r = \text{const}$ — плоскости, перпендикулярные к оси Oz.

Координатные линии:
линии r — лучи, перпендикулярные к оси Oz;
линии φ — окружности с центром на оси Oz в плоскостях, перпендикулярных к оси Oz;
линии z — прямые, параллельные оси Oz.
Формулы преобразования координат:

$$x = r\cos\varphi, \quad y = r\sin\varphi, \quad z = z.$$

Поэтому

$$H_1 = H_r = \sqrt{\left(\frac{\partial x}{\partial r}\right)^2 + \left(\frac{\partial y}{\partial r}\right)^2 + \left(\frac{\partial z}{\partial r}\right)^2} = \sqrt{\cos^2\varphi + \sin^2\varphi} = 1;$$

$$H_2 = H_\varphi = \sqrt{\left(\frac{\partial x}{\partial \varphi}\right)^2 + \left(\frac{\partial y}{\partial \varphi}\right)^2 + \left(\frac{\partial z}{\partial \varphi}\right)^2} = \sqrt{r^2\sin^2\varphi + r^2\cos^2\varphi} = r,$$

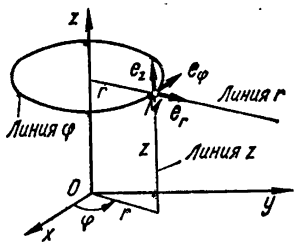

Рис. 82

$$H_3 = H_z = \sqrt{\left(\frac{\partial x}{\partial z}\right)^2 + \left(\frac{\partial y}{\partial z}\right)^2 + \left(\frac{\partial z}{\partial z}\right)^2} = 1.$$

По формулам (16), (17), (18) и (19) получим:

$$\operatorname{grad} U = \frac{\partial U}{\partial r}\boldsymbol{e}_r + \frac{1}{r}\frac{\partial U}{\partial \varphi}\boldsymbol{e}_\varphi + \frac{\partial U}{\partial z}\boldsymbol{e}_z,$$

$$\operatorname{div}\boldsymbol{a} = \frac{1}{r}\left(\frac{\partial (ra_r)}{\partial r} + \frac{\partial a_\varphi}{\partial \varphi} + \frac{\partial (ra_z)}{dz}\right),$$

$$\Delta U = \frac{1}{r}\left(\frac{\partial}{\partial r}\left(r\frac{\partial U}{\partial r}\right) + \frac{\partial}{\partial \varphi}\left(\frac{1}{r}\cdot\frac{\partial U}{\partial \varphi}\right) + \frac{\partial}{\partial z}\left(r\frac{\partial U}{\partial z}\right)\right),$$

$$\operatorname{rot}\boldsymbol{a} = \left(\frac{1}{r}\cdot\frac{\partial a_z}{\partial \varphi} - \frac{\partial a_\varphi}{\partial z}\right)\boldsymbol{e}_r + \left(\frac{\partial a_r}{\partial z} - \frac{\partial a_z}{\partial r}\right)\boldsymbol{e}_\varphi +$$

$$+ \frac{1}{r}\left(\frac{\partial (ra_\varphi)}{\partial r} - \frac{\partial a_r}{\partial \varphi}\right)\boldsymbol{e}_z.$$

Здесь a_r, a_φ, a_z — проекции вектора $\boldsymbol{a}$ на направления $\boldsymbol{e}_r$, $\boldsymbol{e}_\varphi$, $\boldsymbol{e}_z$ соответственно.

Заметим, что цилиндрические координаты используются в теоретической электротехнике при исследовании явления поверхностного эффекта (оттеснения переменного тока от центра к внешней поверхности) в цилиндрическом проводе.

Сферические координаты. Здесь положение точки M определяется координатами ρ, θ и φ (рис. 83). При этом $0 \leqslant \rho < +\infty$, $0 \leqslant \theta \leqslant \pi$, $0 \leqslant \varphi < 2\pi$.

Формулы преобразования координат:

$$x = \rho\sin\theta\cos\varphi,$$
$$y = \rho\sin\theta\sin\varphi,$$
$$z = \rho\cos\theta.$$

Координатные поверхности:
$\rho = \text{const}$ — сферы с центром O;
$\theta = \text{const}$ — конусы с осью Oz;
$\varphi = \text{const}$ — полуплоскости, ограниченные осью Oz.
Координатные линии:
линии ρ — лучи, выходящие из полюса;
линии θ — дуги окружностей с центром O;
линии φ — окружности с центром на оси Oz.

Поэтому

$$H_1 = H_\rho = \sqrt{\left(\frac{\partial x}{\partial \rho}\right)^2 + \left(\frac{\partial y}{\partial \rho}\right)^2 + \left(\frac{\partial z}{\partial \rho}\right)^2} =$$

$$= \sqrt{\sin^2\theta\cos^2\varphi + \sin^2\theta\sin^2\varphi + \cos^2\theta} = 1,$$

$$H_2 = H_\theta = \rho, \quad H_3 = H_\varphi = \rho\sin\theta.$$

Тогда

$$\operatorname{grad} U = \frac{\partial U}{\partial \rho}\cdot \boldsymbol{e}_\rho + \frac{1}{\rho}\cdot\frac{\partial U}{\partial \theta}\boldsymbol{e}_\theta + \frac{1}{\rho\sin\theta}\times$$

$$\times \frac{\partial U}{\partial \varphi}\cdot \boldsymbol{e}_\varphi, \quad \operatorname{div} \boldsymbol{a} = \frac{1}{\rho^2\sin\theta}\left(\sin\theta\times\right.$$

$$\left.\frac{\partial(\rho^2 a_\rho)}{\partial \rho} + \rho\cdot\frac{\partial(\sin\theta\cdot a_\theta)}{\partial\theta} + \rho\cdot\frac{\partial a_\varphi}{\partial\varphi}\right),$$

$$\Delta U = \frac{1}{\rho^2\sin\theta}\left(\sin\theta\cdot\frac{\partial}{\partial\rho}\left(\rho^2\cdot\frac{\partial U}{\partial\rho}\right) + \right.$$

$$\left. + \frac{\partial}{\partial\theta}\left(\sin\theta\cdot\frac{\partial U}{\partial\theta}\right) + \frac{1}{\sin\theta}\cdot\frac{\partial^2 U}{\partial\varphi^2}\right),$$

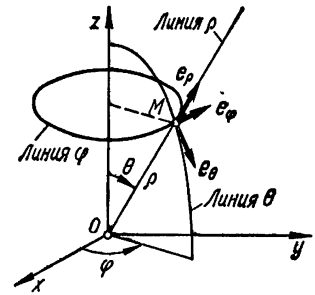

Рис. 83

$$\operatorname{rot} \boldsymbol{a} = \frac{1}{\rho\sin\theta}\left(\frac{\partial(\sin\theta a_\varphi)}{\partial\theta} - \frac{\partial a_\theta}{\partial\varphi}\right)\boldsymbol{e}_\rho + \frac{1}{\rho}\left(\frac{1}{\sin\theta}\cdot\frac{\partial a_\rho}{\partial\varphi} - \right.$$

$$\left. - \frac{\partial(\rho a_\varphi)}{\partial\rho}\right)\boldsymbol{e}_\theta + \frac{1}{\rho}\left(\frac{\partial(\rho a_\theta)}{\partial\rho} - \frac{\partial a_\rho}{\partial\theta}\right)\boldsymbol{e}_\varphi.$$

В электротехнике и радиотехнике сферические координаты используются при рассмотрении вопросов излучения и распространения электромагнитных волн.

ГЛАВА VI

ОСНОВЫ ЛИНЕЙНОЙ АЛГЕБРЫ И ТЕОРИИ МАТРИЦ

Многие явления в природе, физике и технике протекают в пространстве и во времени, изменяясь так, что результат действия каждого отдельного фактора пропорционален его величине, а общий результат действий равен сумме результатов действий отдельных факторов. Такие явления математически описываются

линейными функциями (функциями, зависящими лишь от первых степеней своих аргументов) и их совокупностями.

Учение о линейных функциях, их совокупностях и преобразованиях над ними составляет предмет и существо линейной алгебры. Линейные однородные функции (без свободных членов) описываются системой коэффициентов, совокупность которых задается (записывается) в виде прямоугольной таблицы — матрицы. Теория матриц является важной неотъемлемой частью линейной алгебры, играя роль аппарата при постановке и решении ее задач.

Линейная алгебра широко используется не только во многих разделах математики (аналитической геометрии, теории обыкновенных дифференциальных уравнений, математического анализа и др.) и теоретической физики, но и при решении различных инженерных задач. Наибольшее применение она получила в механике и теоретической электротехнике, в квантовой механике и в теории многосвязанных и многомерных систем автоматического регулирования.

§ 32. ЛИНЕЙНЫЕ ФУНКЦИИ И ЛИНЕЙНЫЕ ПРЕОБРАЗОВАНИЯ

Линейные функции. *Линейной функцией одной независимой переменной называется функция вида* $l(x) = ax + b$. Графиком такой функции является прямая линия на плоскости, ее основные свойства таковы:

приращение линейной функции пропорционально приращению независимой переменной:

$$\Delta l(x) = a\Delta x \quad (a = \text{const}),$$

причем дифференциал функции

$$dl(x) = a\Delta x \quad (\text{т. е. } \Delta l(x) = dl(x));$$

все гладкие (непрерывно дифференцируемые) функции $f(x)$ в первом приближении линейны, так как

$$\Delta f(x) \approx df(x) = f'(x)\,\Delta x.$$

Линейной функцией многих независимых переменных называется функция вида

$$y = a_1x_1 + a_2x_2 + \cdots + a_nx_n + b,$$

где все $a_k = \text{const}$ $(k = 1, 2, 3, \ldots, n)$, $b = \text{const}$.

Если $b = 0$, то линейная функция называется однородной линейной функцией, или линейной формой (под линейной формой еще понимают однородную линейную комбинацию векторов (см. с. 218)).

Основные свойства линейной функции y многих переменных состоят в следующем:

приращение линейной функции y, вычисленное в предположении, что изменяется только одна независимая переменная, про-

порционально приращению этой переменной:

$$\Delta_{x_k} y = a\Delta x_k \quad (k = 1, 2, 3, \ldots, n);$$

приращение линейной функции y в предположении изменения всех независимых переменных равно линейной комбинации приращений этих переменных, причем, если

$$y = \sum_{k=1}^{n} a_k x_k, \text{ то } \Delta y = \sum_{k=1}^{n} a_k \Delta x_k;$$

первое приближение дифференцируемой функции $f(x_1, x_2, \ldots, x_n)$ многих переменных совпадает с линейной функцией, так как

$$\Delta f(x_1, x_2, \ldots, x_n) \approx df(x_1, x_2, \ldots, x_n) = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k} \Delta x_k.$$

Линейные преобразования. *Совокупность (система) линейных форм вида*

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n, \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ y_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{aligned}$$

называется линейным преобразованием переменных x_i $(i = 1, 2, 3, \ldots, n)$ к переменным y_k $(k = 1, 2, 3, \ldots, m)$.

§ 33. МАТРИЦЫ

Линейное преобразование (вне связи с обозначением переменных) полностью характеризуется своими коэффициентами, которые принято записывать с помощью таблицы, ограниченной скобками или двумя парами вертикальных линий:

$$\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \text{ или } \begin{Vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{Vmatrix}.$$

Такая таблица называется матрицей линейного преобразования. В матрице линейного преобразования коэффициенты линейных форм записываются в том порядке, в каком они расположены в линейном преобразовании. Изучение линейного преобразования всегда приводится к изучению его матрицы. Выполнение различных математических операций над линейными формами и линейными преобразованиями весьма громоздко; для этого нужно удерживать в памяти специальные строгие правила. С целью облегчения такой процедуры используется математический образ линейного преобразования — матрица, — в котором исключены переменные. После введения этого образа действия над линейными формами и преобразованиями могут быть заменены более простыми

в исполнении действиями над соответствующими матрицами. Так возникает алгебра матриц. При этом, подобно тому, как в результате действий над линейными преобразованиями вновь получаются линейные преобразования, так в результате действий над матрицами снова получаются матрицы.

Основные определения. *Матрицей (прямоугольной) называется прямоугольная таблица, составленная из чисел (функций).* Числа (функции), из которых составлена матрица, называются ее элементами. Горизонтальные ряды элементов называются строками матрицы, вертикальные ряды — столбцами матрицы.

В общем случае каждый элемент матрицы обозначается одной буквой с двумя индексами; первый индекс указывает номер строки (сверху вниз), а второй — номер столбца (слева направо), на пересечении которых стоит элемент матрицы. Так, например, элемент a_{ik} матрицы стоит на пересечении i-й строки и k-го столбца.

Если в матрице m строк и n столбцов, говорят, что это матрица типа $m \times n$ (размера mn).

Всю матрицу (при необходимости сокращенной записи) принято обозначать одной заглавной буквой, соответствующей малым буквам, которыми обозначены ее элементы, или с помощью символа $[a_{ij}]_{mn}$. Нижний индекс у квадратных скобок указывает тип матрицы, но записывается не всегда. Так,

$$\left\| \begin{matrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{matrix} \right\| = A = [a_{ij}]_{mn} = [a_{ij}].$$

В линейной алгебре важное место занимают следующие частные формы и типы матриц:

Матрица-строка — матрица, состоящая только из одной строки (матрица типа $1 \times n$):

$$A = \| a_1 a_2 \dots a_n \|.$$

Матрица-столбец — матрица, состоящая только из одного столбца (матрица типа $m \times 1$):

$$B = \left\| \begin{matrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{matrix} \right\|.$$

Матрица, состоящая из одного элемента, отождествляется с этим элементом.

Квадратная матрица[1] — матрица, количество строк которой равно количеству столбцов (это количество называется по-

[1] Не следует путать квадратную матрицу с определителем. Матрица — это упорядоченная таблица чисел, а определитель — это число, которое вычисляется по некоторому закону.

рядком квадратной матрицы) —

$$A = \begin{Vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{Vmatrix} = [a_{ij}]_{nn}.$$

С квадратной матрицей обычно связывается определитель, составленный из элементов матрицы, взятых в том же порядке; его называют определителем матрицы и обозначают символом «det A» или Δ (иногда Δ_A).

Так, если

$$A = \begin{Vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{Vmatrix}, \text{ то } \det A = \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} = \Delta_A.$$

Квадратная матрица, определитель которой равен нулю, называется вырожденной (или особенной) матрицей [1]

Прямоугольную матрицу называют ступенчатой, если первая ее строка содержит хотя бы один, не равный нулю элемент, и в каждой строке, начиная со второй, ее первый отличный от нуля элемент расположен правее первого ненулевого элемента в предыдущей строке. (Это определение, естественно, предполагает, что число строк ступенчатой матрицы не больше, чем число ее столбцов).

Квадратная ступенчатая матрица называется треугольной матрицей. Частным случаем треугольной матрицы является диагональная матрица — квадратная матрица, все элементы которой, кроме стоящих на главной диагонали [2], равны нулю:

$$A = \begin{Vmatrix} \alpha_1 & 00 & \dots & 0 \\ 0 & \alpha_2 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 00 & \dots & \alpha_n \end{Vmatrix} = \operatorname{diag}(\alpha_i).$$

Если $\alpha_1 = \alpha_2 = \dots = \alpha_n = \alpha$, то диагональная матрица называется скалярной

$$A = \begin{Vmatrix} \alpha & 0 & \dots & 0 \\ 0 & \alpha & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & \alpha \end{Vmatrix} = \operatorname{diag}(\alpha).$$

В частном случае, когда все $\alpha_i = 1$ ($i = 1, 2, 3, \dots, n$) скалярная матрица называется единичной и обозначается

---

[1] Если $\det A \neq 0$,— матрица невырожденная (неособенная).

[2] Для квадратной матрицы термин «главная диагональ», а также термины «минор» и «алгебраическое дополнение», которые встретятся ниже, совпадают с терминами, применяемыми в теории определителей.

буквой E:

$$E = \begin{Vmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{Vmatrix}.$$

Скалярной матрице всегда соответствует пропорциональное преобразование:

$$\begin{aligned} y_1 &= \alpha x_1 \\ y_2 &= \alpha x_2 \\ y_3 &= \alpha x_3 \\ &\dots\dots\dots\dots \\ y_n &= \alpha x_n, \end{aligned}$$

единичной матрице — тождественное преобразование:

$$\begin{aligned} y_1 &= x_1 \\ y_2 &= x_2 \\ y_3 &= x_3 \\ &\dots\dots \\ y_n &= x_n. \end{aligned}$$

Матрица, все элементы которой нули, называется нулевой и обозначается 0 (или 0_{mn}):

$$0 = \begin{Vmatrix} 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{Vmatrix}.$$

Сравнение двух матриц. Две матрицы A и B называются равными в том и только в том случае, когда они одинаковых размеров и каждый элемент одной матрицы совпадает с соответствующим элементом другой матрицы, т. е. для двух матриц

$$A = \begin{Vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{Vmatrix} \text{ и } B = \begin{Vmatrix} b_{11} & \dots & b_{1p} \\ \vdots & & \vdots \\ b_{q1} & \dots & b_{qp} \end{Vmatrix}.$$

устанавливается равенство $A = B$ только в случае одновременного выполнения условий $p = n$, $q = m$, $a_{ik} = b_{ik}$ ($i = 1, 2, \dots, m$; $k = 1, 2, \dots, n$).

Транспонирование матрицы. Транспонирование матрицы — это замена строк матрицы ее столбцами без изменения последовательности их. Транспонированная матрица обозначается штрихом. Если исходная матрица

$$A = \begin{Vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{Vmatrix},$$

то

$$A' = \begin{Vmatrix} a_{11} & \dots & a_{m1} \\ \vdots & & \vdots \\ a_{1n} & \dots & a_{mn} \end{Vmatrix}.$$

Очевидно:

1. Дважды транспонированная матрица совпадает с исходной матрицей $A'' = (A')' = A$.

2. Если матрица A квадратная, то $\det A' = \det A$.

Матрица называется симметрической, если она совпадает со своей транспонированной, т. е. если $A' = A$. В таком случае матрица будет квадратной и элементы, симметричные относительно главной ее диагонали, равны между собой: $a_{ij} = a_{ji}$.

Если $A' = -A$, то матрица называется кососимметрической.

Ранг матрицы и его вычисление. До введения понятия ранга отметим, что минором k-го порядка матрицы любого размера $m \times n$ называется определитель, полученный из элементов матрицы, стоящих на пересечении произвольных k строк и k столбцов и записанных в том же порядке, что и в матрице.

Например, для матрицы

$$A = \begin{Vmatrix} 3 & 4 & 2 & -1 \\ 5 & 3 & 4 & 2 \\ 1 & 7 & -3 & 9 \end{Vmatrix}$$

можно составить четыре минора третьего порядка

$$\Delta_1 = \begin{vmatrix} 3 & 4 & 2 \\ 5 & 3 & 4 \\ 1 & 7 & -3 \end{vmatrix}, \quad \Delta_2 = \begin{Vmatrix} 3 & 4 & -1 \\ 5 & 3 & 2 \\ 1 & 7 & 9 \end{Vmatrix},$$

$$\Delta_3 = \begin{vmatrix} 3 & 2 & -1 \\ 5 & 4 & 2 \\ 1 & -3 & 9 \end{vmatrix}, \quad \Delta_4 = \begin{vmatrix} 4 & 2 & -1 \\ 3 & 4 & 2 \\ 7 & -3 & 9 \end{vmatrix}$$

и 18 миноров второго порядка

$$\begin{vmatrix} 3 & 4 \\ 5 & 3 \end{vmatrix}, \quad \begin{vmatrix} 3 & 2 \\ 5 & 4 \end{vmatrix}, \quad \begin{vmatrix} 4 & 2 \\ 3 & 4 \end{vmatrix}, \quad \begin{vmatrix} 3 & 4 \\ 1 & 7 \end{vmatrix}, \quad \begin{vmatrix} 2 & -1 \\ -3 & 9 \end{vmatrix}$$

и другие.

Минор первого порядка отождествляется с понятием соответствующего элемента матрицы. Среди всех миноров матрицы могут быть равные нулю и отличные от нуля.

Определение. *Число r такое, что все миноры порядка $k \geqslant r + 1$ равны нулю, а среди миноров каждого порядка $k \leqslant r$ есть ненулевые, называется рангом матрицы.* (Можно также сказать, что *рангом матрицы называется наивысший порядок минора матрицы, отличного от нуля).*

Ранг матрицы играет чрезвычайно важную роль во всех аналитических приложениях матриц. Для вычисления ранга матрицы существует несколько приемов. Рассмотрим некоторые из них.

а) **Вычисление миноров.** Самым примитивным приемом определения ранга матрицы является непосредственное вычисление всех миноров. Здесь можно использовать два варианта — две различные последовательности вычислений.

Во-первых, можно вычислять миноры, начиная со второго порядка, и, как только найден ненулевой минор, переходить к вычислению миноров третьего порядка — до первого ненулевого минора, затем переходить к вычислению миноров четвертого порядка и т. д., до тех пор, пока не окажется, что все миноры некоторого порядка $r + 1$ равны нулю. Ранг матрицы будет равен r.

Во-вторых, можно начинать вычисление всех миноров высших порядков и последовательно понижать порядок миноров, пока не будет найден ненулевой минор. Его порядок будет равен рангу матрицы.

Нетрудно установить, что вычисление всех миноров слишком трудоемкий процесс, поэтому на практике он почти не применяется.

б) **Способ окаймления миноров.** Среди миноров второго порядка выбирают ненулевой минор. Затем вычисляют миноры третьего порядка, окаймляющие (целиком содержащие) выбранный ненулевой минор. Обнаружив среди них ненулевой минор, переходят к вычислению окаймляющих его миноров четвертого порядка и т. д. Порядок последнего из найденных ненулевых миноров оказывается равным рангу матрицы.

Пример. Определить ранг матрицы

$$A = \left\|\begin{array}{cccc} 1 & 1 & -1 & -2 & -2 \\ 2 & 3 & -2 & -5 & -4 \\ 1 & -1 & -1 & 0 & -2 \\ 1 & -2 & -1 & 1 & -2 \end{array}\right\|.$$

Р е ш е н и е. Сразу замечаем, что минор второго порядка, расположенный в левом верхнем углу матрицы (он отмечен пунктирными линиями), отличен от нуля:

$$\begin{vmatrix} 1 & 1 \\ 2 & 3 \end{vmatrix} = 1 \neq 0.$$

Вычисляем теперь окаймляющий его минор третьего порядка (он также отмечен пунктирными линиями):

$$\begin{vmatrix} 1 & 1 & -1 \\ 2 & 3 & -2 \\ 1 & -1 & -1 \end{vmatrix} = 0.$$

Можно убедиться в том, что все остальные миноры третьего порядка, а потом и миноры четвертого порядка равны нулю. Следовательно, ранг матрицы $r = 2$.

в) **Преобразование к ступенчатому виду.** Прежде всего отметим, что ступенчатая матрица имеет ранг l, равный количеству ее ненулевых строк. Действительно, в такой матрице может быть построен хотя бы один минор треугольного вида порядка l. Как известно,

определитель такого вида будет равен произведению элементов определителя, стоящих на главной диагонали, т. е. будет отличен от нуля.

С помощью некоторых преобразований, не изменяющих ранга матрицы и называемых элементарными, любая ненулевая матрица может быть приведена к ступенчатому виду. Это следующие преобразования:

1) перестановка каких-либо двух строк (или двух столбцов) матрицы;

2) умножение всех элементов какой-либо строки (столбца) матрицы на одно и то же отличное от нуля число;

3) прибавление к элементам какой-либо строки соответствующих элементов другой строки, умноженных на одно и то же отличное от нуля число.

Для доказательства того, что эти действия не изменяют ранга матрицы, заметим, что если указанные элементарные преобразования выполнить над определителем, то его абсолютное значение вовсе не изменится или возрастет в $C \neq 0$ раз. Отличный от нуля определитель при таких преобразованиях никогда не станет равным нулю и, наоборот, нулевой определитель не станет отличным от нуля. Поэтому все миноры матрицы, отличные от нуля, останутся неравными нулю, а все нулевые миноры останутся равными нулю. Это значит, что наивысший порядок минора, отличного от нуля, не изменится — ранг матрицы останется тем же самым.

Важно еще отметить, что ранг матрицы не изменится, если матрицу транспонировать.

Элементарные преобразования изменяют элементы матрицы, получается новая матрица, переход к которой указывается горизонтальной стрелкой.

Из сказанного об элементарных преобразованиях следует, что ранг всякой матрицы равен рангу ступенчатой матрицы, полученной из исходной путем элементарных преобразований.

Пример. Определить ранг матрицы

$$A = \begin{Vmatrix} 2 & 1 & 0 & 1 \\ 3 & -1 & -2 & 3 \\ 4 & 3 & 1 & -2 \\ 9 & 3 & -1 & 2 \\ 1 & 3 & 2 & -1 \end{Vmatrix},$$

приводя ее к ступенчатому виду.

Р е ш е н и е. В данной матрице строк больше, чем столбцов, поэтому транспонируем исходную матрицу и определяем ее ранг.

Имеем:

$$A = \begin{Vmatrix} 2 & 3 & 4 & 9 & 1 \\ 1 & -1 & 3 & 3 & 3 \\ 0 & -2 & 1 & -1 & 2 \\ 1 & 3 & -2 & 2 & -1 \end{Vmatrix} \rightarrow$$

вторую и четвертую строки умножаем на 2 и вычитаем из каждой из них первую строку $\rightarrow$

$$\to \begin{Vmatrix} 2 & 3 & 4 & 9 & 1 \\ 2 & -2 & 6 & 6 & 6 \\ 0 & -2 & 1 & -1 & 2 \\ 2 & 6 & -4 & 4 & -2 \end{Vmatrix} \to \begin{Vmatrix} 2 & 3 & 4 & 9 & 1 \\ 0 & -5 & 2 & -3 & 5 \\ 0 & -2 & 1 & -1 & 2 \\ 0 & 3 & -8 & -5 & -3 \end{Vmatrix} \to$$

$\to$ третью и четвертую строки умножаем на 5 и вычитаем из третьей удвоенную вторую строку, а к четвертой прибавляем утроенную вторую строку $\to$

$$\to \begin{Vmatrix} 2 & 3 & 4 & -9 & 1 \\ 0 & -5 & 2 & -3 & 5 \\ 0 & -10 & 5 & -5 & 10 \\ 0 & 15 & -40 & -25 & -15 \end{Vmatrix} \to \begin{Vmatrix} 2 & 3 & 4 & 9 & 1 \\ 0 & -5 & 2 & -3 & 5 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & -34 & -34 & 0 \end{Vmatrix} \to$$

$\to$ третью строку умножаем на 34 и прибавляем к четвертой $\to$

$$\to \begin{Vmatrix} 2 & 3 & 4 & 9 & 1 \\ 0 & -5 & 2 & -3 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{Vmatrix} \to \begin{Vmatrix} 2 & 1 & 3 & 4 & 9 \\ 0 & 0 & -5 & 2 & -3 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{Vmatrix}.$$

Количество ненулевых строк в полученной ступенчатой матрице равно трем. Следовательно, ранг исходной матрицы $r = 3$.

Заметим, что результат остался бы тем же, если бы матрица A не транспонировалась и нули были созданы не в строках, а в столбцах.

Отметим далее, что, вычисляя ранг матрицы, мы иногда приходим к строке из нулевых элементов. В таком случае, очевидно, каждый элемент строки, из которого затем получился нулевой элемент, был линейной комбинацией соответствующих элементов других строк. Всю строку, состоящую из элементов — линейных комбинаций соответствующих элементов других строк, называют строкой — линейной комбинацией других строк (такая строка линейно зависит от других строк).

Если из матрицы вычеркнуть нулевую строку, то ранг матрицы не изменится. То же самое можно сказать и о линейной комбинации строк: если из матрицы вычеркнуть линейно-зависимую строку, то ранг матрицы не изменится.

Очевидно, все сказанное о линейной зависимости и линейной независимости строк матрицы легко может быть перенесено и на ее столбцы.

Действия над матрицами. Среди операций над матрицами, составляющих алгебру матриц, важнейшими являются операции: умножение на число, сложение, вычитание, умножение матриц друг на друга и обращение матриц (для невырожденных квадратных матриц).

В связи с тем что матрицы вводятся на основе линейных преобразований, действия над матрицами определяются на основании соответствующих действий с линейными преобразованиями.

а) **Умножение матрицы на число.** Если некоторая линейная форма умножается на число, то на это число умножается каждое слагаемое линейной формы. Если на число умножается линейное преобразование, то на это число умножается каждая линейная форма, входящая в преобразование, т. е. каждое слагаемое каждой линейной формы. Поэтому умножение матрицы на число равносильно умножению на это число каждого елемента заданной матрицы:

$$\lambda A = \lambda \begin{Vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{Vmatrix} = \begin{Vmatrix} \lambda a_{11} & \dots & \lambda a_{1n} \\ \vdots & & \vdots \\ \lambda a_{m1} & \dots & \lambda a_{mn} \end{Vmatrix}.$$

Из этого, в частности, можно заключить, что для всякой матрицы A существует противоположная матрица, определяемая условием: $-A = (-1) \cdot A$, и что $\operatorname{diag}(\alpha) = E\alpha$.

б) **Сложение двух матриц.** Пусть даны два линейных преобразования:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n, \\ \dots\dots\dots\dots\dots\dots\dots \\ y_m = a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{cases}$$

с матрицей

$$A = \begin{Vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{Vmatrix}$$

и

$$\begin{cases} z_1 = b_{11}x_1 + b_{12}x_2 + \dots + b_{1n}x_n, \\ \dots\dots\dots\dots\dots\dots\dots \\ z_m = b_{m1}x_1 + b_{m2}x_2 + \dots + b_{mn}x_n \end{cases}$$

с матрицей

$$B = \begin{Vmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \dots & b_{mn} \end{Vmatrix}.$$

Сумм о й двух линейных преобразований называется линейное преобразование, в котором каждая линейная форма является суммой соответствующих линейных форм двух исходных преобразований:

$$\begin{cases} w_1 = y_1 + z_1 = (a_{11} + b_{11})x_1 + \dots + (a_{1n} + b_{1n})x_n, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ w_m = y_m + z_m = (a_{m1} + b_{m1})x_1 + \dots + (a_{mn} + b_{mn})x_n. \end{cases}$$

Очевидно, матрицей этого преобразования является матрица

$$C = \begin{Vmatrix} a_{11} + b_{11} & \dots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \dots & a_{mn} + b_{mn} \end{Vmatrix}.$$

Эта матрица называется с у м м о й двух матриц A и B, что записывается так: $C = A + B$.

Таким образом, *суммой двух матриц называется такая матрица, каждый элемент которой равен сумме соответствующих элементов слагаемых матриц.*

Из такого определения следует, что складывать можно только матрицы одинаковых размеров.

Пример. Пусть

$$A = \begin{Vmatrix} 2 & 4 & 5 \\ 1 & 0 & 3 \end{Vmatrix}, \quad B = \begin{Vmatrix} 1 & 4 & 1 \\ -2 & 3 & -1 \end{Vmatrix}.$$

Найти $A + B$.

Р е ш е н и е.

$$A + B = \begin{Vmatrix} 2+1 & 4+4 & 5+1 \\ 1-2 & 0+3 & 3-1 \end{Vmatrix} = \begin{Vmatrix} 3 & 8 & 6 \\ -1 & 3 & 2 \end{Vmatrix}.$$

Операция сложения матриц, для которых определено это действие, обладает следующими основными свойствами:

1. Сложение матриц подчиняется коммутативному (переместительному) закону:

$$A + B = B + A.$$

2. Сложение матриц подчиняется ассоциативному (сочетательному) закону:

$$(A + B) + C = A + (B + C).$$

3. Для любых двух матриц A и B одинаковых типов существует одна и только одна матрица M того же типа такая, что $A + M = B$, т. е. выполняется обратная операция — вычитание:

$$M = A - B = A + (-B).$$

4. Для любых двух матриц одинаковых типов и произвольного числа λ справедливо: $\lambda(A + B) = \lambda A + \lambda B$.

5. Для любых двух чисел μ и λ и матрицы A справедливо:

$$(\mu + \lambda) A = \mu A + \lambda A.$$

6. $(A + B)' = A' + B'$, т. е. транспонированная матрица суммы двух матриц равна сумме соответствующих транспонированных матриц.

в) **Умножение двух матриц.** Умножение двух линейных преобразований выполняется последовательным применением этих преобразований. Так, если к преобразованию

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2, \\ y_2 = a_{21}x_1 + a_{22}x_2, \end{cases}$$

с матрицей

$$A = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

применить преобразование

$$\begin{cases} z_1 = b_{11}y_1 + b_{12}y_2, \\ z_2 = b_{21}y_1 + b_{22}y_2 \end{cases}$$

с матрицей

$$B = \begin{Vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{Vmatrix}.$$

т. е. z_1 и z_2 выразить через x_1 и x_2, то будет выполнено умножение второго преобразования на первое. Результатом такого последовательного применения преобразований (произведением преобразований) будет:

$$z_1 = (b_{11}a_{11} + b_{12}a_{21})\,x_1 + (b_{11}a_{21} + b_{12}a_{22})\,z_2,$$
$$z_2 = (b_{21}a_{11} + b_{22}a_{21})\,x_1 + (b_{21}a_{12} + b_{22}a_{22})\,x_2.$$

Матрица произведения преобразований принимается в качестве произведения матриц преобразований, причем справа записывается матрица первоначально выполненного преобразования:

$$B \cdot A = \begin{Vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{Vmatrix} \begin{Vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{Vmatrix} = \begin{Vmatrix} b_{11}a_{11} + b_{12}a_{21} & b_{11}a_{12} + b_{12}a_{22} \\ b_{21}a_{11} + b_{22}a_{21} & b_{21}a_{12} + b_{22}a_{22} \end{Vmatrix}.$$

Общее правило умножения матриц сводится к следующему:

Каждый элемент (считая слева направо) определенной строки матрицы, стоящей слева, умножается на соответствующий (считая сверху вниз) элемент некоторого столбца матрицы, стоящей справа; полученные произведения складываются; их сумма является элементом матрицы-произведения, стоящим на пересечении строки с номером, равным номеру строки в левом множителе, и столбца с номером, равным номеру столбца в правом множителе. (Правило — «с т р о к а н а с т о л б е ц»).

Таким образом, если $BA = C$, то элементом c_{ik} матрицы произведения является

$$c_{ik} = \sum_{j=1}^{n} b_{ij}a_{jk},$$

где n — количество столбцов в левой матрице (строк — в правой).

Из такого произведения двух матриц можно получить следующее заключение о свойствах умножения матриц:

1. Умножение двух матриц возможно лишь в том случае, когда количество столбцов левой матрицы равно количеству строк правой матрицы [1], при этом матрица-произведение имеет столько строк, сколько строк в левой матрице, и столько столбцов, сколько их в правой матрице.

2. Умножение двух матриц в общем случае неперестановочно, т. е. $AB \neq BA$; матрицы, умножение которых перестановочно, выделены в особый класс коммутативных матриц.

[1] Такие матрицы называются к о н ф о р м н ы м и.

3. Умножение матриц подчиняется сочетательному закону: $(AB)\,C = A\,(BC)$; $\lambda\,(AB) = (\lambda A)\,B = A\,(\lambda B)$, где λ — число.

4. Умножение двух матриц подчиняется распределительному относительно сложения (и вычитания) закону:

$A\,(B + C) = AB + AC$ (соответственно $A\,(B - C) = AB - AC$).

5. Транспонированная матрица произведения равна произведению транспонированных матриц-сомножителей, взятых в обратном порядке: $(AB)' = B'A'$.

Действительно, в матрице $(AB)'$ элемент, стоящий на пересечении i-й строки и j-го столбца, равен элементу матрицы AB, стоящему на пересечении j-й строки и i-го столбца, т. е. сумме $\sum_{k=1}^{n} a_{jk}b_{ki}$. Но это выражение — сумма произведений элементов i-й строки матрицы B' на соответствующие элементы j-го столбца матрицы A', т. е. равно общему элементу матрицы $B'A'$.

Очевидно, произведение $C = AA'$ каждой матрицы на свою транспонированную есть матрица симметрическая, так как

$$C' = (AA')' = (A')'\,A' = AA' = C.$$

Пример.

$$AA' = \begin{Vmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{Vmatrix} \begin{Vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{Vmatrix} = \begin{Vmatrix} 1\cdot 1+4\cdot 4 & 1\cdot 2+4\cdot 5 & 1\cdot 3+4\cdot 6 \\ 2\cdot 1+5\cdot 4 & 2\cdot 2+5\cdot 5 & 2\cdot 3+5\cdot 6 \\ 3\cdot 1+6\cdot 4 & 3\cdot 2+6\cdot 5 & 3\cdot 3+6\cdot 6 \end{Vmatrix} =$$

$$= \begin{Vmatrix} 17 & 22 & 27 \\ 22 & 29 & 36 \\ 27 & 36 & 45 \end{Vmatrix}.$$

6. Среди матриц существуют «делители нуля», т. е. существуют такие пары ненулевых матриц, произведение которых равно нулевой матрице.

Пример.

$$A = \begin{Vmatrix} 4 & -8 \\ -2 & 4 \end{Vmatrix}, \quad B = \begin{Vmatrix} 3 & 6 \\ -6 & -12 \end{Vmatrix},$$

но

$$BA = \begin{Vmatrix} 3 & 6 \\ -6 & -12 \end{Vmatrix} \begin{Vmatrix} 4 & -8 \\ -2 & 4 \end{Vmatrix} = \begin{Vmatrix} 3\cdot 4-6\cdot 2 & 3\cdot(-8)+6\cdot 4 \\ -6\cdot 4+12\cdot 2 & 6\cdot 8-12\cdot 4 \end{Vmatrix} =$$

$$= \begin{Vmatrix} 0 & 0 \\ 0 & 0 \end{Vmatrix} = 0.$$

Нетрудно установить, что для равенства нулю матричного произведения необходимо, чтобы обе перемножаемые матрицы были вырожденными. Но это условие не является достаточным для получения нулевого произведения. Действительно, в приведенном примере

$BA = 0$, однако

$$AB = \begin{Vmatrix} 4 & -8 \\ -2 & 4 \end{Vmatrix} \begin{Vmatrix} 3 & 6 \\ -6 & -12 \end{Vmatrix} = \begin{Vmatrix} 60 & 120 \\ -30 & -60 \end{Vmatrix} \neq 0,$$

а матрицы перемножались те же (хотя и в обратном порядке).

г) **Обращение матриц.** Для каждой квадратной невырожденной матрицы A существует обратная матрица A^{-1} такая, что $A^{-1}A = AA^{-1} = E$.

Вырожденная матрица обратной матрицы не имеет.

Процесс отыскания обратной матрицы называется обращением матрицы.

Чтобы найти обратную матрицу A^{-1} для матрицы A, сперва отыскивают присоединенную (или взаимную) матрицу A^*. С этой целью для каждого элемента a_{ik} исходной квадратной матрицы A составляют соответствующее ему алгебраическое дополнение A_{ik} и вписывают его на место этого элемента. Полученную так матрицу из алгебраических дополнений транспонируют. Таким образом, если

$$A = \begin{Vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{Vmatrix},$$

то

$$A^* = \begin{Vmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \dots & \dots & \dots & \dots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{Vmatrix}.$$

Обратная матрица A^{-1} — это результат деления присоединенной матрицы A^* на определитель исходной матрицы:

$$A^{-1} = \frac{A^*}{\det A},$$

$$A^{-1} = \begin{Vmatrix} \frac{A_{11}}{\Delta} & \frac{A_{21}}{\Delta} & \dots & \frac{A_{n1}}{\Delta} \\ \dots & \dots & \dots & \dots \\ \frac{A_{1n}}{\Delta} & \frac{A_{2n}}{\Delta} & \dots & \frac{A_{nn}}{\Delta} \end{Vmatrix},$$

где $\Delta = \det A$ и A_{ij} = алгебраическое дополнение элемента a_{ij}.

Пример. Дана матрица

$$A = \begin{Vmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 1 & 1 & 1 \end{Vmatrix}.$$

Найти обратную матрицу.

Р е ш е н и е. Вычисляем значения всех алгебраических дополнений элементов заданной матрицы A и ее $\det A$:

$$A_{11} = -2, \quad A_{21} = -1, \quad A_{31} = 5; \quad A_{12} = 1, \quad A_{22} = 0, \quad A_{32} = -1;$$
$$A_{13} = 1, \quad A_{23} = 1, \quad A_{33} = -3;$$
$$\Delta = \det A = \begin{vmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 1 & 1 & 1 \end{vmatrix} = 1 + 6 + 2 - 1 - 4 - 3 = -1;$$

В результате получаем

$$A^* = \begin{Vmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{Vmatrix} = \begin{Vmatrix} -2 & -1 & 5 \\ 1 & 0 & -1 \\ 1 & 1 & -3 \end{Vmatrix}.$$

А потому

$$A^{-1} = \begin{Vmatrix} \frac{A_{11}}{\Delta} & \frac{A_{21}}{\Delta} & \frac{A_{31}}{\Delta} \\ \frac{A_{12}}{\Delta} & \frac{A_{22}}{\Delta} & \frac{A_{32}}{\Delta} \\ \frac{A_{13}}{\Delta} & \frac{A_{23}}{\Delta} & \frac{A_{33}}{\Delta} \end{Vmatrix} = \begin{Vmatrix} -2 & -1 & 5 \\ 1 & 0 & -1 \\ 1 & 1 & -3 \end{Vmatrix}.$$

Нетрудно установить, что обратная матрица соответствует обратному линейному преобразованию. Действительно, рассматривая линейное преобразование с матрицей A, состоящее из n линейных форм, содержащих n переменных $x_1, x_2, \ldots, x_n$

$$\begin{cases} y_1 = a_{11}x_1 + \ldots + a_{1n}x_n, \\ \ldots\ldots\ldots\ldots\ldots\ldots, \\ y_n = a_{n1}x_1 + \ldots + a_{nn}x_n, \end{cases}$$

как систему линейных уравнений относительно $x_1, x_2, \ldots, x_n$, мы приходим (при условии, что определитель системы отличен от нуля, а значит, матрица заданного преобразования невырожденная) к решению системы:

$$x_1 = \frac{1}{\Delta}\begin{vmatrix} y_1 & a_{12} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots & \ldots \\ y_n & a_{n2} & \ldots & a_{nn} \end{vmatrix} = \frac{1}{\Delta}\left(\begin{vmatrix} a_{22} & \ldots & a_{2n} \\ \vdots & & \vdots \\ a_{n2} & \ldots & a_{nn} \end{vmatrix} y_1 + \right.$$

$$+ \begin{vmatrix} a_{12} & \ldots & a_{1n} \\ a_{32} & \ldots & a_{3n} \\ \ldots & \ldots & \ldots \\ a_{n2} & \ldots & a_{nn} \end{vmatrix} y_2 + \ldots + \left. \begin{vmatrix} a_{12} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{(n-1)\,2} & \ldots & a_{(n-1)\,n} \end{vmatrix} y_n \right) =$$

$$= \frac{1}{\Delta}(A_{11}y_1 + A_{21}y_2 + \ldots + A_{n1}y_n),$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$x_n = \frac{1}{\Delta}(A_{1n}y_1 + A_{2n}y_2 + \ldots + A_{nn}y_n).$$

Матрицей этого последнего преобразования и есть матрица A^{-1}.

Если обратная матрица A^{-1} совпадает с транспонированной матрицей, т. е. $A^{-1} = A'$ и $AA' = E$, то матрица A называется ортогональной.

Можно установить следующие факты:

определитель ортогональной матрицы равен ± 1.

матрица, обратная ортогональной, — также ортогональная матрица;

произведение двух ортогональных матриц есть матрица ортогональная.

Заметим, что матрица вида $A^{-1}BA$, где A — произвольная матрица необходимого размера, называется матрицей, подобной матрице B.

Наличие обратной матрицы позволяет ввести для невырожденных квадратных матриц операцию деления, равносильную умножению на обратную матрицу, и соотвественно различать деление слева и деление справа.

Так, если $A = BX$, то $X = B^{-1}A = B\backslash A$ — деление слева, если $A = XB$, то $X = AB^{-1} = A/B$ — деление справа.

Обобщение в случае комплексных чисел. В практике вычислений, особенно в электротехнике при расчете четырехполюсников и сложных электрических цепей переменного тока, приходится иметь дело с матрицами, элементами которых являются комплексные числа. Среди таких матриц важное место занимают комплексно-сопряженные матрицы, унитарные матрицы, эрмитовы и эрмитово-сопряженные матрицы.

Если каждый элемент произвольной матрицы A с комплексными членами заменить комплексно-сопряженным числом, то полученная матрица $\bar{A}$ называется комплексно-сопряженной с A. Для вещественных матриц $A = \bar{A}$.

Унитарной матрицей называется матрица A, для которой справедливо соотношение $AA' = E$.

Матрица, обратная унитарной, а также произведение двух унитарных матриц — это тоже матрицы унитарные.

Легко установить, что упомянутые выше ортогональные матрицы являются частным случаем унитарных матриц (когда элементы унитарных матриц суть вещественные числа).

Эрмитовой называется квадратная матрица, в которой элементы, симметричные главной диагонали, суть комплексно-сопряженные числа, элементы же главной диагонали — обязательно вещественные числа. Например,

$$\begin{Vmatrix} 1 & 1+i & 0 & 3+4i \\ 1-i & 2 & -2i & 5 \\ 0 & 2i & 3 & 2-i \\ 3-4i & 5 & 2+i & 4 \end{Vmatrix}.$$

Вещественная эрмитова матрица всегда симметрическая матрица.

Эрмитово-сопряженная матрица — это матрица A^+, полученная из исходной квадратной матрицы A путем транспонирования и замены каждого элемента соответствующим комплексно-сопряженным числом. Так, если исходная матрица

$$A = \begin{Vmatrix} 1+2i & -4 & 2-3i \\ 1 & 3 & -i \\ 4+i & 2 & i \end{Vmatrix},$$

то ее эрмитово-сопряженная матрица

$$A^+ = \begin{Vmatrix} 1-2i & 1 & 4-i \\ -4 & 3 & 2 \\ 2+3i & i & -i \end{Vmatrix}.$$

Нетрудно заметить, что всякая эрмитова матрица совпадает со своей эрмитово-сопряженной матрицей.

§ 34. ЛИНЕЙНЫЕ ВЕКТОРНЫЕ ПРОСТРАНСТВА

При изучении многих физических явлений и процессов оказывается недостаточным использовать понятие геометрического вектора, а также основанного на этом понятии математического аппарата векторного исчисления. Например, довольно часто требуется знать кроме геометрических координат точки еще несколько физических величин в ней (температуру, плотность, электропроводимость, магнитную и электрическую проницаемость вещества и др.). В таких случаях прибегают к понятиям, применяемым в линейной алгебре и теории матриц, «многомерное пространство» и «многомерный вектор».

В линейной алгебре понятия вектора и векторного пространства широко обобщаются. Многомерное линейное пространство и многомерный вектор здесь вводятся аксиоматически и являются абстрактными понятиями. Природа абстрактного вектора в границах его определения в принципе может быть произвольна. Однако само понятие многомерного вектора тесно связано с понятием линейной формы, а линейное преобразование векторного пространства — с матрицей. Обычные геометрический вектор и трехмерное пространство, в котором мы живем, лишь частный случай многомерных векторов и многомерных пространств.

Основные определения.

Определение. *Линейным векторным пространством*[1] *(элементы которого именуют «векторами») называется некоторое множество элементов, в котором по определенным правилам установлены операция сложения и операция умножения элемента на число (скаляр).*

[1] Коротко: *линейным* пространством или *векторным* пространством.

Целесообразно заметить, что в данном определении совершенно не отражено, в чем состоит операция сложения двух элементов и в чем операция умножения элемента на число. Это дает возможность в каждом конкретном пространстве по-своему определить указанные операции.

Если число вещественное, то векторное пространство вещественное. Если число комплексное, то и векторное пространство — комплексное.

Условимся обозначать: линейное векторное пространство — через R, L и т. п. (прописной буквой); его элементы — через x, y, z, ..., w; числа — через α, β, γ, ..., μ, ... и a, b, c, ...; сумму двух элементов x и y — через $x + y$; произведение элемента x на число λ — через λx.

Для того чтобы множество R было линейным векторным пространством, введенные операции сложения элементов и умножения элемента на число должны удовлетворять следующим условиям (аксиомам):

а) Если $x \in R$, $y \in R$, то $(x + y) \in R$ и $(\lambda x) \in R$.

б) Для любых x, y, z из множества R:

1) $x + y = y + x$;

2) $(x + y) + z = x + (y + z)$;

3) существует элемент $\mathbf{0}$ такой, что $x + \mathbf{0} = x$ (этот элемент $\mathbf{0}$ называется «нуль-вектор»);

4) для каждого x существует элемент y такой, что $x + y = \mathbf{0}$ (y в этом случае называют противоположным элементом для x).

в) Для любых x и y из множества R и для любых чисел α и β:

1) $1 \cdot x = x$;

2) $\alpha(\beta x) = (\alpha\beta)\, x = \alpha\beta x$;

3) $(\alpha + \beta)\, x = \alpha x + \beta x$;

4) $\alpha(x + y) = \alpha x + \alpha y$.

Из сформулированных условий (аксиом), определяющих линейное векторное пространство, следуют две теоремы единственности:

Теорема 1. *В любом линейном векторном пространстве существует единственный нуль.*

Теорема 2. *В любом линейном векторном пространстве для каждого элемента существует единственный противоположный элемент.*

Доказательство этих теорем осуществляется от противного. Выполним его для теоремы 1.

Предположим, что существуют два нулевых вектора $\mathbf{0}_1$ и $\mathbf{0}_2$ (существование хотя бы одного нуля утверждается аксиомой б), 3)). Тогда, считая в аксиоме б), 3), что $x = \mathbf{0}_1$, $\mathbf{0} = \mathbf{0}_2$, получим (так как $x + \mathbf{0} = x$), что $\mathbf{0}_1 + \mathbf{0}_2 = \mathbf{0}_1$.

Теперь, полагая $x = \mathbf{0}_2$, а $\mathbf{0} = \mathbf{0}_1$, найдем, что $\mathbf{0}_2 + \mathbf{0}_1 = \mathbf{0}_2$. Сравнивая полученные равенства (аксиома б), 1)), приходим к заключению, что $\mathbf{0}_1 = \mathbf{0}_2$.

Аналогично доказывается теорема 2.

Из приведенных аксиом также следуют такие утверждения:

1. Для всякого элемента $x \in R$ справедливо: $0 \cdot x = 0$ (слева 0 — число, справа **0** — нуль-вектор).

2. Для всякого элемента $x \in R$ противоположным служит элемент $y = (-1)x$, (который обычно обозначают «$-x$»).

Примеры линейных векторных пространств. а) Пространство V_3 — обычное трехмерное векторное пространство, изучаемое в аналитической геометрии. Сложение векторов выполняется по правилу параллелограмма, умножение вектора на число — это растяжение (сжатие) вектора.

б) Пространство T_n. Элемент этого пространства — совокупность n вещественных чисел: $x = (\xi_1, \xi_2, \xi_3, \ldots, \xi_n)$.

Действия сложения и умножения вектора на число выполняются по таким же правилам, как и с векторами в аналитической геометрии:

$$(\xi_1, \xi_2, \ldots, \xi_n) + (\eta_1, \eta_2, \ldots, \eta_n) = (\xi_1 + \eta_1, \xi_2 + \eta_2, \ldots, \xi_n + \eta_n);$$

$$\lambda(\xi_1, \xi_2, \ldots, \xi_n) = (\lambda\xi_1, \lambda\xi_2, \ldots, \lambda\xi_n).$$

в) Пространство $C(a, b)$. Элемент этого пространства — любая непрерывная функция $x = x(t)$, определенная для всех $t \in [a; b]$. Действия $x_1(t) + x_2(t)$ и $\lambda x(t)$ определяются по правилам математического анализа.

г) Совокупность всех матриц одинакового типа mn. Действия сложения и умножения на число определяются по правилам действий над матрицами:

$$\|a_{ik}\| + \|b_{ik}\| = \|a_{ik} + b_{ik}\|;$$

$$\lambda\|a_{ik}\| = \|\lambda a_{ik}\|.$$

Нетрудно убедиться в том, что все аксиомы, определяющие линейное векторное пространство, во всех приведенных выше примерах выполняются.

Линейная зависимость и линейная независимость векторов. Базис векторного пространства. Пусть $x_1, x_2, \ldots, x_n$ — векторы пространства R, а $C_1, C_2, \ldots, C_n$ — числа.

Тогда вектор $y = C_1x_1 + C_2x_2 + \ldots + C_nx_n$ называется линейной комбинацией векторов $x_1, x_2, \ldots, x_n$.

Если не все $C_i = 0$ ($i = 1, 2, 3, \ldots, n$), но $y = \mathbf{0}$, то векторы $x_1, x_2, \ldots, x_n$ называются линейно-зависимыми.

Если $y = \mathbf{0}$ только тогда, когда $C_1 = C_2 = \ldots = C_n = 0$, то векторы $x_1, x_2, \ldots, x_n$ называются линейно-независимыми.

Напомним, что при изучении линейных дифференциальных уравнений было введено понятие линейной независимости решений дифференциального уравнения. Нетрудно заметить, что при рассмотрении решений уравнения как векторов, принадлежащих пространству C, понятия линейной зависимости для векторов и функ-

ций — решений линейного дифференциального уравнения полностью согласуются. Совершенно аналогичная картина для понятий линейной независимости функций и векторов пространства R. Например, в пространстве V_3 векторы $\boldsymbol{a}_1 = (1;\ 1;\ 1)$, $\boldsymbol{a}_2 = (1;\ 2;\ 3)$, $\boldsymbol{a}_3 = (1;\ 3;\ 2)$ линейно-независимы, так как нет таких C_1, C_2, C_3, чтобы $C_1\boldsymbol{a}_1 + C_2\boldsymbol{a}_2 + C_3\boldsymbol{a}_3 = \boldsymbol{0}$ (C_i не равны нулю одновременно). А три вектора

$$\boldsymbol{b}_1 = (1;\ 1;\ 1),\quad \boldsymbol{b}_2 = (2;\ 2;\ 2),\quad \boldsymbol{b}_3 = (1;\ 2;\ 3)$$

линейно-зависимы, так как $C_1\boldsymbol{b}_1 + C_2\boldsymbol{b}_2 + C_3\boldsymbol{b}_3 = \boldsymbol{0}$, если положить $C_1 = -2C_2$, $C_3 = 0$, C_2 — произвольное отличное от нуля число.

Покажем на примере, как исследуется система векторов на линейную независимость.

Пример. Выяснить, является ли система векторов

$$\boldsymbol{a}_1 = (2;\ -3;\ 1),\quad \boldsymbol{a}_2 = (3;\ -1;\ 5),\quad \boldsymbol{a}_3 = (1;\ -4;\ 3)$$

линейно-зависимой или линейно-независимой.

Р е ш е н и е. Предположим, что система линейно-зависима, т. е. существуют C_1, C_2 и C_3, не все равные нулю, такие, что

$$C_1\boldsymbol{a}_1 + C_2\boldsymbol{a}_2 + C_3\boldsymbol{a}_3 = \boldsymbol{0}.$$

Используя определение действий над векторами и основные аксиомы, получим:

$$C_1\boldsymbol{a}_1 + C_2\boldsymbol{a}_2 + C_3\boldsymbol{a}_3 = C_1(2;\ -3;\ 1) + C_2(3;\ -1;\ 5) + C_3(1;\ -4;\ 3) =$$

$$= (2C_1 + 3C_2 + C_3;\ -3C_1 - C_2 - 4C_3;\ C_1 + 5C_2 + 3C_3) = 0.$$

Следовательно,

$$\begin{cases} 2C_1 + 3C_2 + C_3 = 0, \\ -3C_1 - C_2 - 4C_3 = 0, \\ C_1 + 5C_2 + 3C_3 = 0. \end{cases}$$

Исследуя и решая эту однородную систему линейных уравнений, найдем C_1, C_2, C_3. Определитель системы уравнений:

$$\Delta = \begin{vmatrix} 2 & 3 & 1 \\ -3 & -1 & -4 \\ 1 & 5 & 3 \end{vmatrix} = 35 \neq 0.$$

Система имеет единственное нулевое решение $C_1 = C_2 = C_3 = 0$, следовательно, заданные векторы линейно-независимы.

Использованный здесь прием неудобен в применении. Из примера видно, что нет необходимости всякий раз выписывать все действия над векторами, достаточно сразу записать определитель из координат векторов и выяснить, будет ли он равен нулю или отличен от нуля.

Пример. Исследовать линейную независимость векторов

$$\boldsymbol{a}_1 = (1;\ 2;\ -1;\ 4),\quad \boldsymbol{a}_2 = (3;\ 0;\ 4;\ 2),$$
$$\boldsymbol{a}_3 = (-1;\ 4;\ -6;\ 6),\quad \boldsymbol{a}_4 = (2;\ 3;\ -1;\ 5).$$

Р е ш е н и е. Составляем определитель из координат векторов и вычисляем его:

$$\Delta = \begin{vmatrix} 1 & 2 & -1 & 4 \\ 3 & 0 & 4 & 2 \\ -1 & 4 & -6 & 6 \\ 2 & 3 & -1 & 5 \end{vmatrix} = 0.$$

Так как этот определитель равен нулю, то система векторов линейно-зависима.

Теорема. *Если векторы x_1, x_2, ..., x_n линейно-зависимы, то хотя бы один из них является линейной комбинацией остальных.*

Действительно, если векторы x_1, x_2, ..., x_n линейно-зависимы, то можно записать: $C_1x_1 + C_2x_2 + \ldots + C_nx_n = 0$, где не все C_i ($i = 1, 2, 3, \ldots, n$) одновременно равны нулю, т. е. хотя бы одна из постоянных C_i отлична от нуля. Без ограничения общности можно считать, что $C_1 \neq 0$. Тогда

$$C_1x_1 = -C_2x_2 - C_3x_3 - \cdots - C_nx_n.$$

Поделив на C_1 и полагая

$$-\frac{C_2}{C_1} = \alpha, \quad -\frac{C_3}{C_1} = \beta, \ \ldots, \ -\frac{C_n}{C_1} = \lambda,$$

получим

$$x_1 = \alpha x_2 + \beta x_3 + \cdots + \lambda x_n.$$

Теорема доказана. Совершенно аналогично доказывается и обратная теорема. *Векторы, один из которых есть линейная комбинация остальных, линейно-зависимы.*

Во многих случаях в векторном пространстве R количество линейно-независимых векторов оказывается ограниченным, что целиком определяется природой самого векторного пространства R.

Максимальное число линейно-независимых векторов называется р а з м е р н о с т ь ю векторного пространства.

Так, на плоскости можно выбрать два линейно-независимых вектора, а каждый третий уже будет линейно-зависимым. Следовательно, плоскость — это двумерное векторное пространство.

Точно так же в пространстве V_3 можно выбрать три линейно-независимых вектора, а каждый четвертый будет линейно-зависимым. Поэтому пространство V_3 трехмерно.

Следует заметить, что существуют и бесконечномерные пространства. Например, многочлены от одной переменной образуют бесконечномерное пространство, так как выбранные в нем векторы $1, x, x^2, \ldots, x^N$ линейно-независимы при любом, даже как угодно большом N, т. е. в этом пространстве можно выбрать неограниченное множество линейно-независимых векторов.

О п р е д е л е н и е. *Любые n линейно-независимых векторов $e_1, e_2, e_3, \ldots, e_n$ n-мерного векторного пространства образуют б а з и с этого векторного пространства (векторы $e_1, e_2, \ldots, e_n$ часто называют базисными векторами).*

Такой термин введен вследствие того, что любые n линейно-независимых векторов n-мерного векторного пространства служат базой (основной) для каждого $n+1$-го вектора $\boldsymbol{a}$ (что следует из доказанной выше теоремы), поскольку вектор $\boldsymbol{a}$ n-мерного пространства представляет собой линейную комбинацию n линейно-независимых векторов $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$:

$$\boldsymbol{a} = \alpha_1 \boldsymbol{e}_1 + \alpha_2 \boldsymbol{e}_2 + \cdots + \alpha_n \boldsymbol{e}_n.$$

Такое представление вектора $\boldsymbol{a}$ с помощью векторов $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ единственно и называется разложением вектора $\boldsymbol{a}$ по базису $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \ldots, \boldsymbol{e}_n$.

Числа $\alpha_1, \alpha_2, \ldots, \alpha_n$ называются координатами вектора $\boldsymbol{a}$ в базисе $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$. Это часто записывают так:

$$\boldsymbol{a} = (\alpha_1, \alpha_2, \ldots, \alpha_n) \text{ или } \boldsymbol{a}(\alpha_1, \alpha_2, \ldots, \alpha_n).$$

Если числа α_i $(i = 1, 2, \ldots, n)$ вещественные, то пространство называется вещественным; если эти числа комплексные, то и пространство соответственно называется комплексным векторным пространством.

Сами базисные векторы можно также разложить по тому же базису. Эти разложения имеют вид:

$$\begin{aligned}
\boldsymbol{e}_1 &= (1;\ 0;\ 0;\ \ldots;\ 0),\\
\boldsymbol{e}_2 &= (0;\ 1;\ 0;\ \ldots;\ 0),\\
\boldsymbol{e}_3 &= (0;\ 0;\ 1;\ \ldots;\ 0),\\
&\cdots\cdots\cdots\\
\boldsymbol{e}_n &= (0;\ 0;\ 0;\ \ldots;\ 1).
\end{aligned}$$

Действительно, например, вектор $\boldsymbol{e}_2$ можно представить, как следующую линейную комбинацию базисных векторов:

$$\boldsymbol{e}_2 = 0 \cdot \boldsymbol{e}_1 + 1 \cdot \boldsymbol{e}_2 + 0 \cdot \boldsymbol{e}_3 + \cdots + 0 \cdot \boldsymbol{e}_n.$$

Запись вектора $\boldsymbol{a} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ есть по существу представление его в виде матрицы-строки. Иногда вектор записывают в виде матрицы-столбца:

$$\boldsymbol{a} = \begin{Vmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{Vmatrix}.$$

Установим теперь связь между системой векторов и прямоугольной матрицей.

Пусть дана система векторов:

$$\begin{aligned}
\boldsymbol{x}_1 &= \alpha_1 \boldsymbol{e}_1 + \alpha_2 \boldsymbol{e}_2 + \cdots + \alpha_n \boldsymbol{e}_n,\\
\boldsymbol{x}_2 &= \beta_1 \boldsymbol{e}_1 + \beta_2 \boldsymbol{e}_2 + \cdots + \beta_n \boldsymbol{e}_n,\\
&\cdots\cdots\cdots\cdots\cdots\\
\boldsymbol{x}_m &= \mu_1 \boldsymbol{e}_1 + \mu_2 \boldsymbol{e}_2 + \cdots + \mu_n \boldsymbol{e}_n.
\end{aligned}$$

Каждый вектор этой системы можно представить, как произведение двух матриц. Так, например,

$$x_1 = \|\alpha_1, \alpha_2 \ldots \alpha_n\| \left\| \begin{matrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{matrix} \right\|.$$

Рассматривая матрицу-столбец системы векторов

$$X = \left\| \begin{matrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{matrix} \right\|,$$

можно установить, что

$$X = \left\| \begin{matrix} \alpha_1 \\ \beta_1 \\ \vdots \\ \mu_1 \end{matrix} \right\| \cdot e_1 + \left\| \begin{matrix} \alpha_2 \\ \beta_2 \\ \vdots \\ \mu_2 \end{matrix} \right\| \cdot e_2 + \cdots + \left\| \begin{matrix} \alpha_n \\ \beta_n \\ \vdots \\ \mu_n \end{matrix} \right\| \cdot e_n$$

или

$$X = \left\| \begin{matrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{matrix} \right\| = \left\| \begin{matrix} \alpha_1 & \alpha_2 & \ldots & \alpha_n \\ \beta_1 & \beta_2 & \ldots & \beta_n \\ \cdot & \cdot & \cdot & \cdot \\ \mu_1 & \mu_2 & \ldots & \mu_n \end{matrix} \right\| \left\| \begin{matrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{matrix} \right\|.$$

Мы приходим к выводу, что матрица

$$A = \left\| \begin{matrix} \alpha_1 & \alpha_2 & \ldots & \alpha_n \\ \beta_1 & \beta_2 & \ldots & \beta_n \\ \cdot & \cdot & \cdot & \cdot \\ \mu_1 & \mu_2 & \ldots & \mu_n \end{matrix} \right\|,$$

составленная из координат векторов заданной системы, полностью характеризует эту систему векторов.

Из этого следует, что число линейно-независимых векторов в заданной системе равно числу линейно-независимых строк (столбцов) матрицы, оно равно рангу матрицы A. Если в матрице A есть линейно-зависимые строки, то каждой такой строке соответствует линейно-зависимый вектор из системы X.

Поэтому, в частности, помимо указанного выше способа для исследования линейной независимости системы векторов можно воспользоваться и таким приемом.

Из координат векторов заданной системы составляют матрицу, причем координаты каждого вектора записывают в строку. После этого определяют ранг матрицы системы векторов (его еще коротко называют «ранг системы векторов»), используя способ преобразования к ступенчатой матрице (без ее транспонирования). Если ранг r

матрицы равен количеству векторов, то система векторов линейно-независима. Если ранг r матрицы меньше количества векторов,— система линейно-зависима. Более того, оставшиеся ненулевые строки матрицы можно считать соответствующими базисным векторам заданной системы, нулевые строки — векторам, которые линейно выражаются через базисные векторы [1].

Пример. Исследовать линейную независимость системы векторов

$$a_1 = \begin{Vmatrix} 1 \\ 0 \\ 0 \\ -1 \end{Vmatrix}, \quad a_2 = \begin{Vmatrix} 2 \\ 1 \\ 1 \\ 0 \end{Vmatrix}, \quad a_3 = \begin{Vmatrix} 1 \\ 1 \\ 1 \\ 1 \end{Vmatrix}, \quad a_4 = \begin{Vmatrix} 1 \\ 2 \\ 3 \\ 4 \end{Vmatrix}, \quad a_5 = \begin{Vmatrix} 0 \\ 1 \\ 2 \\ 3 \end{Vmatrix},$$

выбрать базис и выразить через него остальные векторы системы.

Р е ш е н и е. Запишем матрицу A, строками которой являются координаты заданных векторов, и вычислим ее ранг:

$$A = \begin{Vmatrix} 1 & 0 & 0 & -1 \\ 2 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \end{Vmatrix} \to \begin{Vmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 2 \\ 0 & 2 & 3 & 5 \\ 0 & 1 & 2 & 3 \end{Vmatrix} \to \begin{Vmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{Vmatrix} \to$$

$$\to \begin{Vmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{Vmatrix}.$$

Мы приходим к заключению, что ранг r матрицы A равен трем ($r = 3$), третья и пятая строки матрицы A линейно-зависимы, следовательно, линейно-зависимы векторы a_3 и a_5, а базис образуют векторы a_1, a_2, a_4.

Итак, можно записать разложения векторов a_3 и a_5 по базису a_1, a_2, a_4:

$$a_3 = \alpha_1 a_1 + \beta_1 a_2 + \gamma_1 a_4,$$

$$a_5 = \alpha_2 a_1 + \beta_2 a_2 + \gamma_2 a_4.$$

Для нахождения коэффициентов α_i, β_i, γ_i ($i = 1, 2$) необходимо решить две системы четырех уравнений с тремя неизвестными, которые следуют из законов выполнения линейных операций над векторами, заданными в координатах (над координатами выполняются те же линейные операции, что и над векторами; два вектора равны между собой тогда и только тогда, когда совпадают их соответствующие координаты). Имеем:

$$\begin{Vmatrix} 1 \\ 1 \\ 1 \\ 1 \end{Vmatrix} = \alpha_1 \begin{Vmatrix} 1 \\ 0 \\ 0 \\ -1 \end{Vmatrix} + \beta_1 \begin{Vmatrix} 2 \\ 1 \\ 1 \\ 0 \end{Vmatrix} + \gamma_1 \begin{Vmatrix} 1 \\ 2 \\ 3 \\ 4 \end{Vmatrix},$$

$$\begin{Vmatrix} 0 \\ 1 \\ 2 \\ 3 \end{Vmatrix} = \alpha_2 \begin{Vmatrix} 1 \\ 0 \\ 0 \\ -1 \end{Vmatrix} + \beta_2 \begin{Vmatrix} 2 \\ 1 \\ 1 \\ 0 \end{Vmatrix} + \gamma_2 \begin{Vmatrix} 1 \\ 2 \\ 3 \\ 4 \end{Vmatrix}.$$

[1] Координаты векторов можно записать столбцами матрицы, тогда все рассуждения надо будет вести о линейной независимости столбцов (а не строк) матрицы.

Откуда

$$\begin{cases} \alpha_1 + 2\beta_1 + \gamma_1 = 1, \\ \beta_1 + 2\gamma_1 = 1, \\ \beta_1 + 3\gamma_1 = 1, \\ -\alpha_1 + 4\gamma_1 = 1, \end{cases} \qquad \begin{cases} \alpha_2 + 2\beta_2 + \gamma_2 = 0, \\ \beta_2 + 2\gamma_2 = 1, \\ \beta_2 + 3\gamma_2 = 2, \\ -\alpha_2 + 4\gamma_2 = 3 \end{cases}$$

Решая каждую из этих систем, находим:

$$\alpha_1 = -1, \quad \beta_1 = 1, \quad \gamma_1 = 0 \quad \text{и} \quad \alpha_2 = 1, \quad \beta_2 = -1, \quad \gamma_2 = 1.$$

Поэтому разложения векторов a_3 и a_5 по базису a_1, a_2, a_4 будут:

$$a_3 = -a_1 + a_2,$$
$$a_5 = a_1 - a_2 + a_4.$$

Линейные подпространства.

а) О п р е д е л е н и е. *Всякое множество L векторов линейного пространства R, выбранных таким образом, что для каждого вектора $x \in L$ справедливо также $kx \in L$ (где k — число), и для каждых двух векторов $x_1, x_2 \in L$ выполняется условие $(x_1 + x_2) \in L$, называется л и н е й н ы м п о д п р о с т р а н с т в о м пространства R.*

Иными словами, если множество L целиком состоит из векторов линейного пространства R, и при этом произведение любого вектора из L на число и сумма любых двух векторов из L — это также векторы, принадлежащие множеству L, то L называется л и н е й н ы м п о д п р о с т р а н с т в о м пространства R. Это записывается так: $L \subset R$.

Линейное подпространство обладает всеми свойствами линейных пространств. Кроме того, линейные подпространства имеют следующие свойства:

1. Всякое линейное соотношение между векторами подпространства L ($L \subset R$) остается справедливым и во всем пространстве R. В частности, если векторы $x_1, x_2, \ldots, x_k$ подпространства L линейно-зависимы, то они линейно-зависимы и во всем пространстве R.

2. Размерность l линейного подпространства L не превосходит размерности n всего пространства R, т. е. $l \leqslant n$.

3. Всякий базис подпространства L с размерностью l можно дополнить до базиса всего пространства R с размерностью n ($L \subset R$, $l \leqslant n$).

б) П р и м е р ы п о д п р о с т р а н с т в.

1. Нуль-вектор пространства R, очевидно, является наименьшим подпространством пространства R.

2. Все пространство R является наибольшим подпространством пространства R.

3. Если L_1 и L_2 два подпространства пространства R, то совокупность векторов, принадлежащих одновременно L_1 и L_2, образует пересечение подпространств L_1 и L_2 (обозначают $L_1 \cap L_2$). Совокупность векторов вида $y + z$, где $y \in L_1$, $z \in L_2$, называется суммой подпространств L_1 и L_2. Пересечение подпространств и их сумма — это подпространства пространства R.

4. В пространстве V_3 совокупность всех векторов, параллельных одной и той же фиксированной плоскости, образует подпространство пространства V_3.

5. Совокупность диагональных матриц порядка n образует подпространство пространства квадратных матриц порядка n.

Линейные оболочки.

а) Определение. *Линейной оболочкой векторов $x_1, x_2, \ldots, x_k$ пространства R называется совокупность всех возможных (конечных) линейных комбинаций этих векторов:*

$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k$, *где* $\alpha_1, \alpha_2 \ldots, \alpha_k$ — *числа.*

Например, если даны два вектора x_1 и x_2 из пространства R, то векторы $2x_1, 3x_1, \ldots, 15x_1, \ldots, mx_1; 2x_2, 3x_2, \ldots, mx_2$ принадлежат линейной оболочке векторов x_1 и x_2; линейной оболочке этих векторов x_1 и x_2 принадлежат также векторы $x_1 + x_2, 2x_1 + x_2, x_1 - 2x_2, 15x_1 - 0{,}3x_2$ и вообще векторы $k_1x_1 + k_2x_2$, где k_1 и k_2 — числа. Таким образом, линейная оболочка векторов x_1 и x_2 состоит из всех возможных линейных комбинаций этих векторов, т. е. из всех векторов, порожденных векторами x_1 и x_2.

Для линейной оболочки векторов выполняются требования, предъявляемые к подпространствам, а потому линейная оболочка системы векторов $x_1, x_2, \ldots, x_k$ из R является подпространством пространства R. Кроме того, всякое подпространство, содержащее векторы $x_1, x_2, \ldots, x_k$, содержит и все их линейные комбинации; поэтому линейная оболочка векторов $x_1, x_2, \ldots, x_k$ — это наименьшее подпространство, содержащее эти векторы.

Линейную оболочку векторов $x_1, x_2, \ldots, x_k$ принято обозначать $L(x_1, x_2, \ldots, x_k)$, а векторы $x_1, x_2, \ldots, x_k$ — называть векторами, порождающими соответствующую линейную оболочку.

б) Примеры линейных оболочек.

1. Линейная оболочка базисных векторов $e_1, e_2, \ldots, e_n$ пространства R: $\alpha_1e_1 + \alpha_2e_2 + \ldots + \alpha_ne_n$, очевидно, совпадает с самим пространством R.

2. Линейная оболочка пары неколлинеарных векторов a и b пространства V_3: $\alpha a + \beta b$ есть плоскость, состоящая из всех векторов, параллельных плоскости векторов a и b.

в) Основные свойства линейных оболочек.

1. Линейная оболочка векторов $x_1, y_1, \ldots$, входящих в состав линейной оболочки $L(x, y, \ldots)$, содержится в этой линейной оболочке.

Действительно, так как векторы $x_1, y_1, \ldots$ принадлежат линейному подпространству $L(x, y, \ldots)$, то все их линейные комбинации также принадлежат подпространству L, а совокупность упомянутых линейных комбинаций и есть линейная оболочка векторов $x_1, y_1, \ldots$, т. е. она содержится в подпространстве $L(x, y, \ldots)$.

2. Исключение из системы векторов $x, y, z, \ldots, w$ вектора, линейно зависящего от остальных векторов системы, не изменяет их линейной оболочки. Этот факт подтверждается следующим рассуждением.

Если, например, $x = \alpha y + \beta z + \ldots + \gamma w$, то $x \in L(y, z, \ldots, w)$. Поэтому (на основании первого свойства линейных оболочек)

$$L(x, y, z, \ldots, w) \subset L(y, z, \ldots, w).$$

Но с другой стороны (на основании того же свойства)

$$L(y, z, \ldots, w) \subset L(x, y, z, \ldots, w).$$

Из этого непосредственно следует $L(x, y, z, \ldots, w) = L(y, z, \ldots, w)$, что и требовалось доказать.

Если среди векторов x, y, ..., w, порождающих линейную оболочку, есть k линейно-независимых векторов x_1, x_2, ..., x_k через которые линейно выражаются все остальные векторы системы x, y, ..., w, то векторы x_1, x_2, ..., x_k образуют базис линейной оболочки, и размерность линейной оболочки, очевидно, равна k.

Линейные формы. Линейные операторы. Выше были введены понятия линейной функции, линейной формы и линейного преобразования. Эти понятия переносятся на векторные пространства, которые используются для описания и изучения реальных физических процессов и явлений, происходящих в пространстве и во времени.

Определение. *Если каждому вектору $x \in R$ поставлено в соответствие некоторое число $f(x)$, причем для любых $x \in R$, $y \in R$ выполняются условия:*

1) $f(x + y) = f(x) + f(y)$;

2) $f(\alpha x) = \alpha f(x)$ *(α — любое число), то числовая функция $f(x)$ векторного аргумента называется линейной формой.*

Два условия, определяющие линейность, могут быть сведены вместе и для двух векторов записаны в виде:

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y),$$

а для n векторов — в виде

$$f(\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n) = \alpha_1 f(x_1) + \alpha_2 f(x_2) + \cdots + \alpha_n f(x_n),$$

или

$$f\left(\sum_{i=1}^{n} \alpha_i x_i\right) = \sum_{i=1}^{n} \alpha_i f(x_i)$$

для любых векторов $x_i \in R$ и для любых чисел α_i ($i = 1, 2, \ldots, n$).

Применим линейную форму $f(x)$ к конкретному вектору

$$x = \sum_{k=1}^{n} \xi_k e_k \quad (e_k \text{ — базисные векторы}).$$

Обозначим $f(e_k) = c_k$. Тогда

$$f(x) = f\left(\sum_{k=1}^{n} \xi_k e_k\right) = \sum_{k=1}^{n} \xi_k f(e_k) = \sum_{k=1}^{n} \xi_k c_k,$$

т. е. значения линейной формы линейно выражаются через координаты вектора.

Для того чтобы упростить изучение физических процессов и явлений при их математическом описании (при рассмотрении векторов и линейных форм) выполняется ряд преобразований.

Определение. *Преобразования, применяемые к векторам, называют операторами (операторы будем обозначать малыми греческими буквами:* σ, τ, ν, μ *и т. п.).*

Оператор σ считается заданным в линейном пространстве, если существует функция, которая каждому вектору $x \in R$ ставит в соответствие некоторый другой вектор $x' \in R$.

Вектор x называется оригиналом, вектор x' — образом (изображением). Это записывают в одной из двух форм:

$$x \to x' \quad \text{или} \quad x' = \sigma x.$$

Если для оператора σ выполняются условия линейности:

1) $\sigma(x_1 + x_2) = \sigma x_1 + \sigma x_2 = x_1' + x_2'$;

2) $\sigma(cx) = c\sigma x = cx'$ (где c — произвольный скаляр, число), то оператор σ называется линейным оператором.

а) Примеры линейных операторов.

1. Нулевой оператор — оператор, ставящий каждому вектору $x \in R$ в соответствие нуль-вектор: $\sigma x = 0$.

2. Единичный (тождественный) оператор — оператор, сохраняющий векторы неизменными (переводящий каждый вектор в самого себя): $\sigma x = x$.

3. Оператор подобия — оператор, переводящий каждый вектор в коллинеарный ему: $\sigma x = \lambda x$ (λ — скаляр).

4. Оператор, переводящий базисные векторы e_k в коллинеарные им:

$$\sigma e_k = \lambda_k e_k \quad (k = 1, 2, \ldots, n; \quad \lambda_k \text{ — число}),$$

а любой другой вектор

$$x = \sum_{k=1}^{n} \xi_k e_k \text{ — в вектор } \sigma x = \sum_{k=1}^{n} \lambda_k \xi_k e_k.$$

Этот оператор называется диагональным.

б) Простейшие свойства линейных операторов.

1. При любом линейном преобразовании нуль-вектору соответствует нуль-вектор.

2. Если $\sigma x = x'$, то $\sigma(-x) = -x'$.

3. Совокупность M всех образов x' векторов $x \in R_n$ есть подпространство пространства R_n: $M \subset R_n$.

Как и в других случаях изучения линейных преобразований, при изучении линейных операторов применяется матричное исчисление.

в) Основные задачи, связанные с применением операторов. Обычно рассматриваются следующие три основные задачи:

1. Установление общего вида линейного оператора при переходе к новому базису.

2. Изучение изменения координат вектора при линейном преобразовании.

3. Изучение изменения матрицы линейного преобразования при переходе к новому базису.

Рассмотрим каждую из этих трех задач:

г) Переход к новому базису. Пусть σ — некоторый оператор, а векторы $e_1, e_2, \ldots, e_n$ образуют базис в пространстве R_n (индексом n отмечена размерность пространства).

Оператор σ переводит каждый базисный вектор e_i в некоторый новый вектор f_i:

$$\sigma e_i = f_i \quad (i = 1, 2, 3, \ldots, n).$$

Каждый из полученных векторов f_i может быть разложен по первоначальному базису:

$$\sigma e_i = \sum_{k=1}^{n} a_{ik} e_k,$$

т. е.

$$\begin{aligned}
f_1 &= \sigma e_1 = a_{11}e_1 + a_{12}e_2 + \cdots + a_{1n}e_n,\\
f_2 &= \sigma e_2 = a_{21}e_1 + a_{22}e_2 + \cdots + a_{2n}e_n,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
f_n &= \sigma e_n = a_{n1}e_1 + a_{n2}e_2 + \cdots + a_{nn}e_n.
\end{aligned}$$

Коэффициенты выписанных формул преобразования базисных векторов определяют матрицу

$$A = \begin{Vmatrix} a_{11} & a_{21} & \ldots & a_{n1} \\ a_{12} & a_{22} & \ldots & a_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ a_{1n} & a_{2n} & \ldots & a_{nn} \end{Vmatrix},$$

которая называется матрицей оператора σ в базисе e.

Столбцы этой матрицы суть координаты векторов $f_1, f_2, \ldots, f_n$.

Такую запись делают для того, чтобы применяя векторы e_k и f_k в виде матриц-столбцов, иметь $f_k = Ae_k$.

Справедлива следующая теорема, которую мы приведем без доказательства.

Теорема. *Существует один и только один линейный оператор σ, переводящий векторы $e_1, e_2, \ldots, e_n$ пространства R_n соответственно в векторы $f_1, f_2, \ldots, f_n$, где $\{e_i\}$ векторы базиса, а $\{f_i\}$ — произвольная система векторов.* (Символом $\{a_k\}$ принято обозначать последовательность (систему) величин a_k, $k \in N$).

Сформулированная теорема позволяет, указав только матрицу, говорить: «задан оператор», не добавляя слов «с матрицей», т. е. отождествлять матрицу и оператор.

На основании приведенной теоремы заключаем, что возможно построение линейного оператора σ, переводящего векторы $\boldsymbol{e}_i$ в векторы $\boldsymbol{f}_i$ ($i = 1, 2, 3, \ldots, n$) по выписанным выше формулам:

$$\sigma\boldsymbol{e}_1 = \boldsymbol{f}_1, \quad \sigma\boldsymbol{e}_2 = \boldsymbol{f}_2, \quad \ldots, \quad \sigma\boldsymbol{e}_n = \boldsymbol{f}_n,$$

причем матрицей оператора σ будет матрица A. Такой оператор единственный.

д) Примеры матриц операторов.

1. Матрица нулевого оператора состоит только из нулей.

2. Матрица тождественного оператора — это единичная матрица

$$E = \begin{Vmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & 1 \end{Vmatrix}.$$

3. Матрица оператора подобия — диагональная матрица, ненулевые (диагональные) элементы которой равны коэффициенту подобия λ:

$$\begin{Vmatrix} \lambda & 0 & \ldots & 0 \\ 0 & \lambda & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & \lambda \end{Vmatrix}.$$

4. Матрица диагонального оператора — диагональная матрица, ненулевые (диагональные) элементы которой равны коэффициентам λ_k:

$$\begin{Vmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & \lambda_n \end{Vmatrix}.$$

5. Матрица A оператора σ, переводящего базисные векторы $\boldsymbol{e}_1 = (1, 0, 0)$, $\boldsymbol{e}_2 = (0, 1, 0)$, $\boldsymbol{e}_3 = (0, 0, 1)$ в векторы соответственно $\boldsymbol{f}_1 = (0, 0, 1)$, $\boldsymbol{f}_2 = (0, 1, 1)$, $\boldsymbol{f}_3 = (1, 1, 1)$, имеет вид

$$A = \begin{Vmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{Vmatrix},$$

так как

$$\boldsymbol{f}_1 = \sigma\boldsymbol{e}_1 = 0 \cdot \boldsymbol{e}_1 + 0 \cdot \boldsymbol{e}_2 + 1 \cdot \boldsymbol{e}_3,$$
$$\boldsymbol{f}_2 = \sigma\boldsymbol{e}_2 = 0 \cdot \boldsymbol{e}_1 + 1 \cdot \boldsymbol{e}_2 + 1 \cdot \boldsymbol{e}_3,$$
$$\boldsymbol{f}_3 = \sigma\boldsymbol{e}_3 = 1 \cdot \boldsymbol{e}_1 + 1 \cdot \boldsymbol{e}_2 + 1 \cdot \boldsymbol{e}_3.$$

е) Изменение координат вектора при воздействии линейного оператора. Зная матрицу A оператора σ, можно представить координаты любого вектора,

преобразованного оператором A в базисе e, через координаты первоначального (преобразуемого) вектора.

Действительно, пусть оператор σ применяется к вектору $x = \sum_{i=1}^{n} \xi_i e_i$ и в результате получается вектор

$$y = \sum_{k=1}^{n} \eta_k e_k, \text{ т. е. } \sigma x = y.$$

Тогда

$$y = \sum_{k=1}^{n} \eta_k e_k = \sigma x = \sigma\left(\sum_{i=1}^{n} \xi_i e_i\right) = \sum_{i=1}^{n} \xi_i \sigma e_i =$$

$$= \sum_{i=1}^{n} \xi_i \sum_{k=1}^{n} a_{ik} e_k = \sum_{k=1}^{n}\left(\sum_{i=1}^{n} \xi_i a_{ik}\right) e_k,$$

где матрицей оператора σ является матрица

$$A = \begin{Vmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{Vmatrix}.$$

Из записанного следует (так как два вектора равны между собой тогда и только тогда, когда совпадают их соответствующие координаты в одном и том же базисе), что

$$\eta_k = \sum_{i=1}^{n} a_{ik}\xi_i \quad (k = 1, 2, 3, \dots, n),$$

или в развернутом виде

$$\begin{cases} \eta_1 = a_{11}\xi_1 + a_{21}\xi_2 + \dots + a_{n1}\xi_n, \\ \eta_2 = a_{12}\xi_1 + a_{22}\xi_2 + \dots + a_{n2}\xi_n, \\ \dots\dots\dots\dots\dots\dots \\ \eta_n = a_{1n}\xi_1 + a_{2n}\xi_2 + \dots + a_{nn}\xi_n. \end{cases}$$

Нетрудно заметить, что матрица коэффициентов при координатах ξ_i ($i = 1, 2, 3, \dots, n$) совпадает с матрицей оператора σ.

Таким образом, становится возможным, применяя оператор σ к любому вектору, определить координаты его образа, полученного в результате воздействия оператора.

Выполненное преобразование (применение оператора σ к вектору x и получение в результате этого вектора y) можно записать в матричной форме, что сокращает вычисления. Для этого необходимо каждый из векторов x и y записать в виде матрицы-столбца:

$$x = \begin{Vmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{Vmatrix}, \qquad y = \begin{Vmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{Vmatrix}$$

и тогда

$$\sigma x = \begin{Vmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{Vmatrix} \begin{Vmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{Vmatrix} =$$

$$= \begin{Vmatrix} a_{11}\xi_1 + a_{21}\xi_2 + \dots + a_{n1}\xi_n \\ a_{12}\xi_1 + a_{22}\xi_2 + \dots + a_{n2}\xi_n \\ \dots \\ a_{1n}\xi_1 + a_{2n}\xi_2 + \dots + a_{nn}\xi_n \end{Vmatrix} = \begin{Vmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{Vmatrix} = y.$$

ж) Изменение матрицы линейного оператора (преобразования) при переходе к новому базису. Пусть в пространстве R_n базисные векторы $\{e_i\}$ переводятся в базисные векторы $\{f_i\}$ с помощью оператора σ, имеющего матрицу

$$A = \begin{Vmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{Vmatrix},$$

т. е.

$$\sigma e_i = f_i \quad (i = 1, 2, 3, \dots, n).$$

Пусть далее в том же пространстве действует оператор β, матрицу которого в базисе $\{e_i\}$ обозначим через B:

$$B = \begin{Vmatrix} b_{11} & b_{21} & \dots & b_{n1} \\ b_{12} & b_{22} & \dots & b_{n2} \\ \dots & \dots & \dots & \dots \\ b_{1n} & b_{2n} & \dots & b_{nn} \end{Vmatrix},$$

а в базисе $\{f_i\}$ — через C:

$$C = \begin{Vmatrix} c_{11} & c_{21} & \dots & c_{n1} \\ c_{12} & c_{22} & \dots & c_{n2} \\ \dots & \dots & \dots & \dots \\ c_{1n} & c_{2n} & \dots & c_{nn} \end{Vmatrix}.$$

Очевидно

$$\beta e_k = \sum_{i=1}^{n} b_{ik} e_i,$$

$$\beta f_k = \sum_{i=1}^{n} c_{ik} f_i \quad (k = 1, 2, 3, \dots, n).$$

Найдем выражение матрицы C через матрицы B и A. Для этого в последнем равенстве заменим векторы f_k через σe_k, а векторы f_i —

через σe_i. Тогда

$$\beta\sigma e_k = \sum_{i=1}^{n} c_{ik}\sigma e_i \quad (k = 1, 2, 3, \ldots, n).$$

Для того чтобы в правой части равенства иметь выражение, зависящее непосредственно от базисных векторов $\{e_i\}$, применим к обеим частям этого равенства обратный оператор σ^{-1} (здесь мы учитываем, что $\sigma^{-1} \cdot \sigma$ — это тождественное преобразование, а обратное преобразование σ^{-1} существует, так как векторы $\{f_i\}$ линейно-независимы, а потому матрица A — невырожденная). Имеем

$$\sigma^{-1}(\beta\sigma e_k) = \sum_{i=1}^{n} \sigma^{-1}(c_{ik}\sigma e_i) = \sum_{i=1}^{n} c_{ik}(\sigma^{-1}\sigma)\, e_i = \sum_{i=1}^{n} c_{ik} e_i.$$

Таким образом, $(\sigma^{-1}\beta\sigma)\, e_k = \sum_{i=1}^{n} c_{ik} e_i$, откуда следует, что матрица преобразования $(\sigma^{-1}\beta\sigma)$ в базисе $\{e_i\}$ совпадает с матрицей C.

Но матрица произведения операторов в одном и том же базисе есть произведение матриц этих операторов в том же базисе. Поэтому матрица преобразования $(\sigma^{-1}\beta\sigma)$ есть матрица $A^{-1}BA$.

Итак, матрица C получена: $C = A^{-1}BA$, где A — есть матрица перехода от базиса $\{e_i\}$ к базису $\{f_i\}$, т. е. матрица B оператора β в базисе $\{e_i\}$ при переходе к новому базису преобразуется в подобную ей матрицу C.

Собственные векторы и собственные значения оператора. Особую роль в изучении векторных пространств играют инвариантные (неизменные) одномерные подпространства оператора. Их также называют инвариантными, или собственными направлениями.

Напомним, что в курсе аналитической геометрии в пространстве V_3 два вектора, параллельных одной и той же прямой, назывались коллинеарными векторами. Этот термин переносится и на пространства большей размерности. На основании сказанного выше, дадим следующее определение:

Определение. *Ненулевой вектор x называется собственным вектором оператора σ, если он этим оператором переводится в коллинеарный вектор λx т. е. $\sigma x = \lambda x$, где λ — некоторое число (скаляр).*

Все такие числа λ называются собственными значениями оператора σ. Например, для нулевого оператора каждый вектор является собственным вектором с собственным значением 0. Для оператора подобия также каждый вектор есть собственный вектор, но с собственным значением λ, а оператор поворота в пространстве V_2 не имеет собственных векторов с вещественными значениями.

Нетрудно установить, что:

1. В n-мерном пространстве линейный оператор не может иметь более чем n собственных векторов с различными собственными значениями.

2. Собственные векторы оператора, соответствующие собственному значению λ, образуют подпространство пространства R_n.

Выясним, как вычислить собственные значения и собственные векторы в n-мерном пространстве R_n.

Пусть векторы $\{e_i\}$ образуют базис пространства R_n, σ — линейный оператор с матрицей A:

$$A = \begin{Vmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{Vmatrix}.$$

Предположим, что вектор $x = \sum_{i=1}^{n} \xi_i e_i$ есть собственный вектор оператора σ, т. е. $\sigma x = \lambda x$, где λ — некоторое собственное значение. Найдем координаты вектора x и собственное значение λ. Для этого применим к вектору x оператор σ. Очевидно,

$$\sigma x = \sum_{i=1}^{n} \xi_i \sigma e_i = \lambda x.$$

Переписывая это соотношение в развернутом виде в координатной форме, имеем

$$\begin{cases} \lambda\xi_1 = a_{11}\xi_1 + a_{21}\xi_2 + \dots + a_{n1}\xi_n, \\ \lambda\xi_2 = a_{12}\xi_1 + a_{22}\xi_2 + \dots + a_{n2}\xi_n, \\ \dots\dots\dots\dots\dots\dots\dots \\ \lambda\xi_n = a_{1n}\xi_1 + a_{2n}\xi_2 + \dots + a_{nn}\xi_n, \end{cases}$$

или

$$\begin{cases} (a_{11} - \lambda)\,\xi_1 + a_{21}\xi_2 + \dots + a_{n1}\xi_n = 0, \\ a_{12}\xi_1 + (a_{22} - \lambda)\,\xi_2 + \dots + a_{n2}\xi_n = 0, \\ \dots\dots\dots\dots\dots\dots\dots \\ a_{1n}\xi_1 + a_{2n}\xi_2 + \dots + (a_{nn} - \lambda)\,\xi_n = 0. \end{cases}$$

Эта система линейных уравнений относительно значений координат $\xi_1, \xi_2, \dots, \xi_n$ однородна, поэтому она допускает ненулевые решения тогда и только тогда, когда ее определитель равен нулю[1], т. е. когда

$$\Delta(\lambda) = \begin{vmatrix} a_{11} - \lambda & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} - \lambda & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{nn} - \lambda \end{vmatrix} = 0.$$

Многочлен, стоящий в левой части этого равенства (и представленный определителем), называется характеристическим многочленом оператора σ, а все равенство (уравнение

[1] Подробнее о разрешимости систем линейных уравнений см. в § 35 данной главы, а также в курсе аналитической геометрии.

относительно λ) — характеристическим уравнением этого оператора.

Из характеристического уравнения оператора σ находятся значения [1]λ, а затем из однородной системы (для каждого значения λ) — соответствующие значения координат вектора x.

Пример. Найти собственные векторы линейного оператора, имеющего в некотором базисе матрицу

$$\begin{Vmatrix} 4 & -5 & 7 \\ 1 & -4 & 9 \\ -4 & 0 & 5 \end{Vmatrix}.$$

Р е ш е н и е. Система уравнений относительно координат собственного вектора $x = \sum_{i=1}^{3} \xi_i e_i$ имеет вид

$$\begin{cases} (4-\lambda)\,\xi_1 - 5\xi_2 + 7\xi_3 = 0, \\ \xi_1 - (4+\lambda)\,\xi_2 + 9\xi_3 = 0, \\ -4\xi_1 + (5-\lambda)\,\xi_3 = 0. \end{cases}$$

Характеристическое уравнение:

$$\begin{vmatrix} 4-\lambda & -5 & 7 \\ 1 & -(4+\lambda) & 9 \\ -4 & 0 & 5-\lambda \end{vmatrix} = 0,$$

или после упрощения: $\lambda^3 - 5\lambda^2 + 17\lambda - 13 = 0$. Его корни: $\lambda_1 = 1$, $\lambda_2 = 2 - 3i$, $\lambda_3 = 2 + 3i$.

В результате получим три самостоятельные системы уравнений относительно ξ_1, ξ_2, ξ_3, каждая — соответственно выбранному собственному значению λ:

1. Для $\lambda_1 = 1$:

$$\begin{cases} 3\xi_1 - 5\xi_2 + 7\xi_3 = 0, \\ \xi_1 - 5\xi_2 + 9\xi_3 = 0, \\ -4\xi_1 + 4\xi_3 = 0. \end{cases}$$

Здесь $\xi_2 = 2\xi_1$, $\xi_3 = \xi_1$. Поэтому собственные векторы, соответствующие собственному значению $\lambda_1 = 1$, имеют вид: C_1 (1; 2; 1).

2. Для $\lambda_2 = 2 - 3i$:

$$\begin{cases} (2+3i)\,\xi_1 - 5\xi_2 + 7\xi_3 = 0, \\ \xi_1 - (6-3i)\,\xi_2 + 9\xi_3 = 0, \\ -4\xi_1 + (3+3i)\,\xi_3 = 0. \end{cases}$$

В этом случае

$$\xi_1 = \frac{3+3i}{4}\,\xi_3, \quad \xi_2 = \frac{5+3i}{4}\,\xi_3$$

и собственные векторы имеют вид: C_2 (3 + 3i; 5 + 3i; 4).

3. Для $\lambda_3 = 2 + 3i$:

$$\begin{cases} (2-3i)\,\xi_1 - 5\xi_2 + 7\xi_3 = 0, \\ \xi_1 - (6+3i)\,\xi_2 + 9\xi_3 = 0, \\ -4\xi_1 + (3-3i)\,\xi_3 = 0. \end{cases}$$

[1] Очевидно, не более чем n различных значений, что подтверждает высказанное выше утверждение.

Теперь

$$\xi_1 = \frac{3-3i}{4}\xi_3, \quad \xi_2 = \frac{5-3i}{4}\xi_3,$$

поэтому собственные векторы будут: $C_3(3-3i;\ 5-3i;\ 4)$.

В заключение этого пункта отметим без доказательства, что собственные значения линейного оператора (а значит, и его собственные векторы) не зависят от выбора базиса. Это связано с тем, что характеристический многочлен оператора остается неизменным в любом базисе.

§ 35. СИСТЕМЫ ЛИНЕЙНЫХ УРАВНЕНИЙ

Пусть дана система линейных алгебраических уравнений:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n = b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n = b_m. \end{cases}$$

Числа $b_1, b_2, b_3, \ldots, b_m$ называются свободными членами уравнений. Если хотя бы один из свободных членов отличен от нуля, система уравнений называется неоднородной системой; если все $b_i = 0$ $(i = 1, 2, 3, \ldots, m)$, то система однородна.

Зная действия над матрицами, можно указанную систему уравнений привести к матричной форме, которая поможет решить систему.

Действительно, записанная система уравнений представляет собой совокупность линейных форм, и, если неизвестные $x_1, x_2, \ldots, x_n$ считать элементами матрицы-столбца X, где

$$X = \begin{Vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Vmatrix},$$

свободные члены $b_1, b_2, \ldots, b_m$ — элементами другой матрицы-столбца B_1, где

$$B_1 = \begin{Vmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{Vmatrix},$$

а коэффициенты при неизвестных $a_{11}, a_{12}, \ldots, a_{mn}$ — элементами прямоугольной матрицы A, где

$$A = \begin{Vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{Vmatrix},$$

то заданную систему уравнений можно переписать в виде одного матричного уравнения $AX = B_1$.

И тогда, если матрица A имеет обратную матрицу A^{-1}, легко получить решение линейной системы $X = A^{-1}B_1$.

Однако такой подход к решению системы линейных уравнений не всегда возможен — матрица A может оказаться вырожденной (а потому не иметь обратной матрицы), возможно, что $m \neq n$, т. е. количество уравнений не равно количеству неизвестных и т. п.

Поэтому решению системы линейных уравнений должно предшествовать предварительное исследование, которое вместе с решением системы подводит нас к следующим основным задачам теории систем линейных уравнений:

1. Исследовать, является ли система уравнений совместной или несовместной (т. е. имеется ли вообще решение системы).

2. Если система совместна, то выяснить, является ли она определенной (т. е. имеется ли единственное решение системы, или таких решений бесконечное количество).

3. Если система совместна и определенна, — найти ее единственное решение.

4. Если система совместна и неопределенна,— описать всю совокупность ее решений.

Перейдем к последовательному рассмотрению этих задач, причем изучим сначала неоднородные системы уравнений.

Для системы m линейных уравнений с n неизвестными

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

составим матрицу A из коэффициентов при неизвестных x_1, x_2, x_3, ..., x_n

$$A = \begin{Vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{Vmatrix}$$

и матрицу B, содержащую помимо всех столбцов матрицы A еще и столбец свободных членов,

$$B = \begin{Vmatrix} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{Vmatrix}.$$

Первая из этих матриц (матрица A) называется матрицей системы, или основной матрицей, вторая (матрица B) — расширенной матрицей.

Исследование совместности системы. Справедливо следующее утверждение (теорема Кронекера — Капелли — критерий совместности системы линейных уравнений):

Теорема. *Система линейных уравнений совместна тогда и только тогда, когда ранг расширенной матрицы равен рангу матрицы системы (этот общий ранг называется рангом системы).*

Для доказательства теоремы Кронекера — Капелли сперва покажем, что система линейных уравнений равносильна одному векторному уравнению. Примем столбцы расширенной матрицы B за векторы (см. с. 211):

$$\boldsymbol{b}_1 = \begin{Vmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{Vmatrix}, \quad \boldsymbol{b}_2 = \begin{Vmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{Vmatrix}, \quad \ldots, \quad \boldsymbol{b}_n = \begin{Vmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{Vmatrix}, \quad \boldsymbol{b} = \begin{Vmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{Vmatrix}.$$

Неизвестные $x_1, x_2, \ldots, x_n$ считаем обычными числами. Тогда вместо заданной системы уравнений получится одно векторное уравнение

$$\boldsymbol{b}_1 x_1 + \boldsymbol{b}_2 x_2 + \cdots + \boldsymbol{b}_n x_n = \boldsymbol{b}.$$

Пусть система чисел $(\beta_1, \beta_2, \ldots, \beta_n)$ — решение записанного векторного уравнения. Подставляя вместо x_k значения β_k $(k = 1, 2, 3, \ldots, n)$, получим тождество. Учитывая, что

$$\boldsymbol{b}_i = (a_{1i}, a_{2i}, \ldots, a_{mi}) \quad (i = 1, 2, 3, \ldots, n),$$
$$\boldsymbol{b} = (b_1, b_2, \ldots, b_m),$$

выполним действия, соответствующие подстановке решения $\{\beta_k\}$ $(k = 1, 2, 3, \ldots, n)$ в векторное уравнение:

$$\boldsymbol{b} = \boldsymbol{b}_1\beta_1 + \boldsymbol{b}_2\beta_2 + \cdots + \boldsymbol{b}_n\beta_n = (a_{11}\beta_1 + a_{12}\beta_2 + \cdots + a_{1n}\beta_n;$$
$$a_{21}\beta_1 + a_{22}\beta_2 + \cdots + a_{2n}\beta_n; \ \ldots; \ a_{m1}\beta_1 +$$
$$+ a_{m2}\beta_2 + \cdots + a_{mn}\beta_n) = (b_1, b_2, \ldots, b_m).$$

Из равенства векторов следует равенство их проекций:

$$\begin{cases} a_{11}\beta_1 + a_{12}\beta_2 + \cdots + a_{1n}\beta_n = b_1, \\ a_{21}\beta_1 + a_{22}\beta_2 + \cdots + a_{2n}\beta_n = b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}\beta_1 + \alpha_{m2}\beta_2 + \cdots + \alpha_{mn}\beta_n = b_m. \end{cases}$$

Мы показали, что система уравнений имеет то же решение, что и векторное уравнение.

Обратное утверждение, что всякое решение системы уравнений есть решение соответствующего векторного уравнения, доказывается в обратном порядке.

Теперь перейдем к доказательству критерия совместности системы уравнений.

1. Пусть система уравнений

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

совместна, т. е. имеет решение. Обозначим его $(\beta_1, \beta_2, \ldots, \beta_n)$.

Это решение будет также решением соответствующего векторного уравнения $\boldsymbol{b}_1x_1 + \boldsymbol{b}_2x_2 + \ldots + \boldsymbol{b}_nx_n = \boldsymbol{b}$:

$$\boldsymbol{b}_1\beta_1 + \boldsymbol{b}_2\beta_2 + \cdots + \boldsymbol{b}_n\beta_n = \boldsymbol{b}.$$

Но последнее равенство указывает на линейную зависимость вектора $\boldsymbol{b}$ от векторов $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n$. А это значит, что, не изменяя ранга матрицы B, из нее можно исключить вертикальный вектор $\boldsymbol{b}$ (это следует из свойств линейных оболочек см. с. 217). Но тогда остается матрица A, поэтому ранг расширенной матрицы B оказывается равным рангу матрицы системы A.

2. Пусть ранг основной матрицы A равен рангу r расширенной матрицы B.

Так как ранг матрицы A равен r, то в этой матрице есть хотя бы один минор D порядка r, отличный от нуля. Этот определитель будет одновременно минором расширенной матрицы B. Без ограничения общности можно считать, что этот минор D расположен, например, в верхнем левом углу каждой из матриц A и B (в противном случае мы переставим уравнения и перенумеруем неизвестные так, чтобы минор D, отличный от нуля, оказался в левом верхнем углу матриц A и B). Тогда можно утверждать, что первые r векторов-строк матрицы B линейно-независимы, а все последующие — линейно-зависимы от них. Отсюда следует, что уравнения, соответствующие этим последующим строкам, являются лишь простыми следствиями первых r уравнений, и можно ограничиться рассмотрением только первых r уравнений,— их решения будут удовлетворять и всем остальным уравнениям системы. Система имеет решение — она совместна.

Таким образом, теорема Кронекера — Капелли полностью доказана.

В дальнейшем условимся обозначать через r_A — ранг матрицы A, через r_B — ранг матрицы B, если $r_A = r_B$, то ранг системы — через r.

Исследование единственности решения и отыскание решения. Очевидно, возможны два случая: $r = n$ или $r < n$.

а) Если $r = n$, т. е. ранг системы уравнений равен числу неизвестных, то, отбросив уравнения-следствия, мы сохраним количество уравнений, равное числу неизвестных. В оставленной так системе уравнений определитель системы будет отличен от нуля. Матрица системы A будет иметь обратную матрицу, и мы можем рассматривать вместо системы линейных уравнений одно матричное уравнение, как было указано в начале данного параграфа: $AX = B_1$.

Решение такого уравнения единственно (так как всякая невырожденная матрица имеет одну-единственную обратную матрицу), а именно: $X = A^{-1}B_1$.

Здесь мы приходим к формулам Крамера, которые являются простым следствием записанного в матричной форме решения системы линейных уравнений.

Пример. Найти решение системы уравнений

$$\begin{cases} x_1 + x_2 + x_3 = 3, \\ 2x_1 + 3x_2 - x_3 = -4, \\ x_1 - x_2 + x_3 = 5, \\ 4x_1 + 3x_2 + x_3 = 4. \end{cases}$$

Р е ш е н и е. Здесь

$$A = \begin{Vmatrix} 1 & 1 & 1 \\ 2 & 3 & -1 \\ 1 & -1 & 1 \\ 4 & 3 & 1 \end{Vmatrix}, \quad B = \begin{Vmatrix} 1 & 1 & 1 & 0 \\ 2 & 3 & -1 & -4 \\ 1 & -1 & 1 & 5 \\ 4 & 3 & 1 & 4 \end{Vmatrix}.$$

Для вычисления рангов r_A и r_B удобно рассматривать сразу расширенную матрицу B, отделив в ней условно столбец свободных членов, и при вычислении рангов элементарные преобразования выполнять только над строками матрицы. Имеем

$$B = \left\| \begin{array}{ccc:c} 1 & 1 & 1 & 3 \\ 2 & 3 & -1 & -4 \\ 1 & -1 & 1 & 5 \\ 4 & 3 & 1 & 4 \end{array} \right\| \to \left\| \begin{array}{ccc:c} 1 & 1 & 1 & 3 \\ 2 & 3 & -1 & -4 \\ 4 & 3 & 1 & 4 \\ 4 & 3 & 1 & 4 \end{array} \right\| \to \left\| \begin{array}{ccc:c} 1 & 1 & 1 & 3 \\ 2 & 3 & -1 & -4 \\ 4 & 3 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{array} \right\| \to$$

$$\to \left\| \begin{array}{ccc:c} 1 & 1 & 1 & 3 \\ 0 & 1 & -3 & -10 \\ 0 & -1 & -3 & -8 \\ 0 & 0 & 0 & 0 \end{array} \right\| \to \left\| \begin{array}{ccc:c} 1 & 1 & 1 & 3 \\ 0 & 1 & -3 & -10 \\ 0 & 0 & -6 & -18 \\ 0 & 0 & 0 & 0 \end{array} \right\|.$$

Отсюда, $r_A = r_B = 3$.

Одно уравнение оказывается следствием трех остальных. Оставляем систему из первых трех уравнений и решаем ее по формулам Крамера:

$$\begin{cases} x_1 + x_2 + x_3 = 3, \\ 2x_1 + 3x_2 - x_3 = -4, \\ x_1 - x_2 + x_3 = 5. \end{cases}$$

Определитель системы

$$\Delta = \begin{vmatrix} 1 & 1 & 1 \\ 2 & 3 & -1 \\ 1 & -1 & 1 \end{vmatrix} = -6.$$

Числители

$$\Delta_{x_1} = \begin{vmatrix} 3 & 1 & 1 \\ -4 & 3 & -1 \\ 5 & -1 & 1 \end{vmatrix} = -6; \quad \Delta_{x_2} = \begin{vmatrix} 1 & 1 & 1 \\ 2 & -4 & -1 \\ 1 & 5 & 1 \end{vmatrix} = 6;$$

$$\Delta_{x_3} = \begin{vmatrix} 1 & 1 & 3 \\ 2 & 3 & -4 \\ 1 & -1 & 5 \end{vmatrix} = -18.$$

По формулам Крамера имеем

$$x_1 = \frac{\Delta_{x_1}}{\Delta} = 1, \quad x_2 = \frac{\Delta_{x_2}}{\Delta} = -1, \quad x_3 = \frac{\Delta_{x_3}}{\Delta} = 3.$$

б) Если $r < n$, т. е. ранг матрицы системы меньше, чем число неизвестных, то мы вновь отбрасываем уравнения-следствия, оставляя лишь r уравнений. Но теперь оказывается, что $n - r$ неизвестных «свободны», т. е. не связаны никакими условиями. Это приводит нас к заключению, что, подставляя вместо каждого свободного неизвестного произвольное значение, мы получим бесчисленное множество систем уравнений, каждая из которых будет иметь единственное решение. Вследствие этого исходная система уравнений имеет бесчисленное множество решений, т. е. неопределенна.

Для решения такой системы мы оставим в левых частях уравнений r неизвестных, а все остальные неизвестные перенесем вправо и, рассматривая их как составные части свободных членов, выразим r неизвестных через $n - r$ свободных неизвестных.

Пример. Исследовать, совместна ли система уравнений, и, если совместна, решить ее:

$$\begin{cases} x_1 - x_2 + 2x_3 - x_4 = 1, \\ x_1 - x_2 + x_3 - x_4 = 0, \\ x_1 - x_2 + 5x_3 - x_4 = 4, \\ x_1 - x_2 + 6x_3 - x_4 = 5. \end{cases}$$

Р е ш е н и е. Определяем ранги r_A и r_B основной матрицы A и расширенной матрицы B:

$$B = \left\|\begin{array}{cccc:c} 1 & -1 & 2 & -1 & 1 \\ 1 & -1 & 1 & -1 & 0 \\ 1 & -1 & 5 & -1 & 4 \\ 1 & -1 & 6 & -1 & 5 \end{array}\right\| \to \left\|\begin{array}{cccc:c} 1 & -1 & 2 & -1 & 1 \\ 1 & -1 & 1 & -1 & 0 \\ 1 & -1 & 5 & -1 & 4 \\ 0 & 0 & 4 & 0 & 4 \end{array}\right\| \to$$

$$\to \left\|\begin{array}{cccc:c} 1 & -1 & 2 & -1 & 1 \\ 1 & -1 & 1 & -1 & 0 \\ 0 & 0 & 3 & 0 & 3 \\ 0 & 0 & 4 & 0 & 4 \end{array}\right\| \to \left\|\begin{array}{cccc:c} 1 & -1 & 2 & -1 & 1 \\ 0 & 0 & -1 & 0 & -1 \\ 0 & 0 & 3 & 0 & 3 \\ 0 & 0 & 4 & 0 & 4 \end{array}\right\| \to$$

$$\to \left\|\begin{array}{cccc:c} 1 & -1 & 2 & -1 & 1 \\ 0 & 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right\|.$$

Отсюда видно, что $r_A = r_B = 2$.

Ранг системы меньше числа неизвестных. Система неопределенна.

Независимыми можно считать первые два уравнения (соответствующие ненулевым строкам матрицы, полученной в результате преобразований при вычислении рангов):

$$\begin{cases} x_1 - x_2 + 2x_3 - x_4 = 1, \\ x_1 - x_2 + x_3 - x_4 = 0. \end{cases}$$

В качестве основных неизвестных выберем x_2 и x_3, а x_1 и x_4 будем считать свободными неизвестными. Тогда

$$\begin{cases} x_2 - 2x_3 = x_1 - x_4 - 1, \\ x_2 - x_3 = x_1 - x_4, \end{cases}$$

а определитель системы отличен от нуля:

$$\Delta = \begin{vmatrix} 1 & -2 \\ 1 & -1 \end{vmatrix} = 1.$$

Числители

$$\Delta_{x_2} = \begin{vmatrix} x_1 - x_4 - 1 & -2 \\ x_1 - x_3 & -1 \end{vmatrix} = x_1 - x_4 + 1,$$

$$\Delta_{x_3} = \begin{vmatrix} 1 & x_1 - x_4 - 1 \\ 1 & x_1 - x_4 \end{vmatrix} = 1.$$

Поэтому $x_2 = x_1 - x_4 + 1$, $x_3 = 1$.

Перейдем к рассмотрению однородных систем уравнений. Нетрудно установить, что однородная система уравнений

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0 \end{cases}$$

всегда совместна, так как ранг матрицы системы всегда в этом случае равен рангу расширенной матрицы (добавление нулевого столбца свободных членов не изменяет ранга матрицы), нулевое решение всегда существует. Поэтому может возникнуть вопрос лишь о единственности решения. И этот вопрос решается, как обычно: если ранг r_A основной матрицы равен количеству неизвестных, то система имеет единственное (нулевое) решение; если ранг r системы меньше, чем число n неизвестных ($r < n$), то система неопределенна (кроме нулевого решения она имеет бесчисленное множество ненулевых решений), и тогда выбирают основные и «свободные» неизвестные. Через последние выражается общее решение системы подобно тому, как это делается для неоднородных систем в случае $r < n$. Однако для неопределенных однородных систем принято записывать так называемые фундаментальную систему решений и нормальную фундаментальную систему решений. Эти понятия основаны на том факте, что множество всех решений произвольной однородной линейной системы уравнений с n неизвестными образует подпространство пространства T_n n-мерных строк чисел $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ — сумма двух решений и произведение решения на число остаются решениями той же системы уравнений.

Если ранг однородной системы уравнений с n неизвестными равен r ($r < n$), то r неизвестных выражаются через остальные $n - r$ линейно-независимых неизвестных; в таком случае любая система из $n-r$ линейно-независимых решений (система $n - r$ линейно-независимых векторов-строк) образует базис пространства всех

решений и поэтому называется фундаментальной системой решений.

Простейший базис всякого пространства состоит из так называемых нормированных векторов. Для рассматриваемого n—r-мерного пространства решений таким простейшим базисом будет $e_1 = (1, 0, 0, \ldots, 0)$, $e_2 = (0, 1, 0, \ldots, 0)$, ..., $e_{n-r} = (0, 0, 0, \ldots, 1)$. Чтобы получить решение, соответствующее вектору e_1, надо в решение вместо свободных неизвестных подставить координаты вектора e_1, аналогично — для вектора e_2 и т. д. Построенная таким образом совокупность решений называется нормальной фундаментальной системой решений.

Если обозначить решения, соответствующие векторам e_1, e_2, ..., e_{n-r} через $x^{(1)}$, $x^{(2)}$, ..., $x^{(n-r)}$, то, по определению базиса, для любого решения будет

$$x = C_1x^{(1)} + C_2x^{(2)} + \cdots + C_{n-r}x^{(n-r)}.$$

Такое выражение называется общим решением однородной системы линейных уравнений.

Пример. Найти нормальную фундаментальную систему решений для системы уравнений

$$\begin{cases} x_1 + x_2 + x_3 + x_4 + x_5 = 0, \\ 3x_1 + 2x_2 + x_3 + x_4 - 3x_5 = 0, \\ x_2 + 2x_3 + 2x_4 + 6x_5 = 0, \\ 5x_1 + 4x_2 + 3x_3 + 3x_4 - x_5 = 0. \end{cases}$$

Решение. Прежде всего определяем ранг системы. Нетрудно установить, что $r = 2$. Следовательно, в заданной системе есть два линейно-независимых уравнения.

Примем в качестве основных первое и третье уравнения, в качестве основных неизвестных — x_1 и x_2 и выразим их через остальные неизвестные, т. е. считаем x_3, x_4, x_5 свободными неизвестными. Имеем

$$\begin{cases} x_1 + x_2 = -x_3 - x_4 - x_5, \\ x_2 = -2x_3 - 2x_4 - 6x_5. \end{cases}$$

Отсюда

$$x_1 = x_3 + x_4 + 5x_5,$$
$$x_2 = -2x_3 - 2x_4 - 6x_5.$$

Нами получена система решений, построенная на выбранных свободных неизвестных; подставляя произвольные значения x_3, x_4, x_5, будем получать соответствующие значения x_1 и x_2.

Таблица 4

x_1	x_2	x_3	x_4	x_5
1	—2	1	0	0
1	—2	0	1	0
5	—6	0	0	1

Положим $x_3 = 1$, $x_4 = x_5 = 0$. Тогда первое решение будет: $x^{(1)} = (1, -2, 1, 0, 0)$.

Аналогично получим два других решения нормальной фундаментальной системы. Их часто сводят в таблицу. Так, для рассматриваемой системы решений имеем табл. 4.

Замечание. Для неоднородной системы уравнений, для которой ранг r меньше количества n неизвестных, можно также построить общее решение.

При этом не трудно доказать, что общее решение неоднородной системы уравнений представляет собой сумму общего решения однородной системы, соответствующей данной неоднородной системе, и какого угодно решения неоднородной системы (сравните это положение с положением, известным из теории линейных дифференциальных уравнений).

§ 36. ПРИМЕРЫ ПРИМЕНЕНИЯ АППАРАТА ЛИНЕЙНОЙ АЛГЕБРЫ И МАТРИЧНОГО ИСЧИСЛЕНИЯ

Рассмотрим некоторые приложения аппарата линейной алгебры и матричного исчисления в теоретической электротехнике, теории автоматического регулирования и в механике.

Применение в электротехнике. В настоящее время имеется стройная теория применения матричного исчисления к исследованию электрических цепей. Проиллюстрируем ее основные исходные положения на примере.

Пусть нам задана принципиальная схема сложной линейной электрической цепи постоянного тока (рис. 84, *а*), сопротивления которой не зависят от величин напряжений и токов.

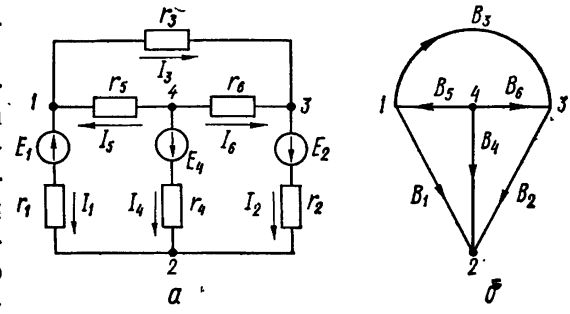

Рис. 84

Тогда ей будет соответствовать скелетная схема цепи (рис. 84, *б*), на которой узлы (полюса) обозначены арабскими цифрами 1, 2, 3, 4, а ветви — прописными буквами русского алфавита с цифровыми индексами: B_1, B_2, B_3, B_4, B_5, B_6. Ветвям скелетной схемы придают направления, совпадающие с произвольно выбранными в начале исследования (расчета) положительными направлениями токов в них (направлениями напряжений на зажимах ветвей).

Если теперь в скелетной схеме цепи выделить некоторую совокупность ветвей так, чтобы она связывала между собой все узлы и при этом не образовывала ни одного замкнутого контура, то получим так называемое дерево схемы. Ветви дерева схемы обычно вычерчивают жирными линиями (рис. 85). Ветви, не входящие в дерево, называют главными ветвями схемы. Для заданной цепи можно получить несколько различных деревьев, из которых берется любое.

Далее определяют число независимых замкнутых контуров n цепи, имеющей p ветвей и $q + 1$ узел, по формуле $n = p - q$. Выбирают эти независимые контуры и произвольно задаются положительными направлениями их обхода (обычно всюду по ходу стрелки часов либо против ее хода). Независимые контуры обозначают буквой K с цифровым индексом (рис. 85).

Если в скелетной схеме цепи провести ряд замкнутых линий так, чтобы каждая из них пересекала только одну из ветвей дерева,

а остальные пересекаемые ветви относились бы к главным ветвям, то в результате получим несколько замкнутых областей схемы, называемых сечениями. Число сечений схемы равно q. Сечения удобно изображать штрих-пунктирными линиями и обозначать буквой C с цифровым индексом. Сечениям также придаются положительные направления, но уже внутрь и вне выделяемой ими замкнутой области (рис. 85).

Имея подготовленную рассмотренным способом скелетную схему заданной электрической цепи, можно описать ее алгебраически единственным образом с помощью одной из двух матриц, составленных из чисел +1, —1 и 0. Одна из них называется матрицей сечений, вторая — матрицей контуров.

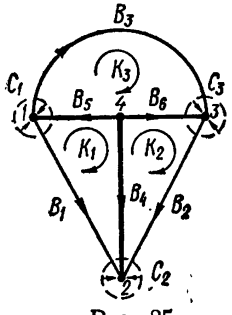

Рис. 85

Матрица сечений выражает взаимно однозначное соответствие сечений и ветвей схемы и определяет связь между напряжениями на зажимах ветвей цепи и потенциалами ее узлов. Она обозначается буквой Π и строится по следующему правилу: строка отводится сечению C_i, столбец — ветви B_j, если направления k-го сечения и m-й ветви, включаемой в это сечение, совпадают, то на пересечении k-й строки и m-го столбца матрицы ставится +1; если эти направления противоположны, то вписывается —1; если k-е сечение не включает m-ю ветвь, то ставится 0.

Для скелетной схемы цепи, приведенной на рис. 85, матрица сечений имеет вид:

$$\Pi = \begin{array}{c|cccccc} & B_1 & B_2 & B_3 & B_4 & B_5 & B_6 \\ \hline C_1 & +1 & 0 & -1 & 0 & +1 & 0 \\ C_2 & -1 & +1 & 0 & +1 & 0 & 0 \\ C_3 & 0 & -1 & +1 & 0 & 0 & +1 \end{array} = \begin{Vmatrix} +1 & 0 & -1 & 0 & +1 & 0 \\ -1 & +1 & 0 & +1 & 0 & 0 \\ 0 & -1 & +1 & 0 & 0 & +1 \end{Vmatrix}.$$

Матрица контуров выражает взаимно однозначное соответствие независимых контуров и ветвей схемы и определяет связь между токами в ветвях цепи и контурными токами в ней (положительные направления которых совпадают с направлениями обхода независимых контуров). Эта матрица обозначается буквой Γ и строится по такому правилу: в ней строка отводится контуру K_i, столбец — ветви B_j; если направления k-го контура и m-й ветви, входящей в этот контур, совпадают, то на пересечении k-й строки и m-го столбца матрицы ставится +1; если эти направления противоположны, то вписывается —1; если k-й контур не проходит через m-ю ветвь, то ставится 0.

Для рассматриваемого примера (рис. 85) матрица контуров имеет вид:

$$\Gamma = \begin{array}{c|cccccc} & B_1 & B_2 & B_3 & B_4 & B_5 & B_6 \\ \hline K_1 & +1 & 0 & 0 & +1 & -1 & 0 \\ K_2 & 0 & +1 & 0 & -1 & 0 & +1 \\ K_3 & 0 & 0 & +1 & 0 & +1 & -1 \end{array} = \left\| \begin{matrix} +1 & 0 & 0 & +1 & -1 & 0 \\ 0 & +1 & 0 & -1 & 0 & +1 \\ 0 & 0 & +1 & 0 & +1 & -1 \end{matrix} \right\|.$$

Если теперь ввести матрицы-столбцы I — токов ветвей и U — напряжений ветвей, то можно получить выражения законов Кирхгофа в матричной форме:

$\Pi I = 0$ — уравнение первого закона Кирхгофа,

$\Gamma U = 0$ — уравнение второго закона Кирхгофа.

Для анализа сложных электрических цепей используются так называемые матрично-векторные параметры схемы, которые для цепей постоянного тока удобно обозначить через r, g, e и i.

Матрица сопротивлений ветвей r является квадратной диагональной матрицей порядка p. Элементы ее главной диагонали равны сопротивлениям r_m соответствующих ветвей схемы цепи, а все остальные элементы равны нулю.

Матрица проводимостей ветвей g — также квадратная диагональная матрица порядка p. Элементы ее главной диагонали равны проводимостям g_m соответствующих ветвей схемы цепи, а все остальные элементы равны нулю. Имеем:

$$r = \left\| \begin{matrix} r_1 & 0 & 0 & \dots & 0 \\ 0 & r_2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & r_p \end{matrix} \right\|, \quad g = \left\| \begin{matrix} g_1 & 0 & 0 & \dots & 0 \\ 0 & g_2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & g_p \end{matrix} \right\|.$$

Так как для цепей постоянного тока $g_m = \frac{1}{r_m}$, то для одной и той же схемы цепи матрицы r и g будут взаимообратными, т. е.

$$r = g^{-1} \quad \text{и} \quad g = r^{-1}.$$

Матрица э. д. с. ветвей e-столбцовая порядка p. Она записывается для цепи, в ветвях которой действуют источники э. д. с.

Матрица задающих токов ветвей i также столбцовая матрица порядка p. Но составляется она для цепи, содержащей в своих ветвях источники тока.

В результате для цепи с источниками э. д. с. система p уравнений относительно напряжений на зажимах всех ветвей в матричной форме будет иметь вид

$$U = rI - e,$$

а система p уравнений относительно токов во всех ветвях цепи с источниками тока — вид

$$I = gU + i.$$

Эти уравнения цепей в совокупности с уравнениями, выражающими законы Кирхгофа, являются основными матричными уравнениями схем электрических цепей. Они позволяют решать сложные задачи анализа электрических цепей с многополюсными элементами.

Пример из автоматики. Уравнения движения линейных автоматических систем выглядят так:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = f_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = f_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = f_n. \end{cases}$$

В матричной форме эта система уравнений запишется в виде:

$$Ax = f,$$

где

$$A = \begin{Vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{Vmatrix}, \quad x = \begin{Vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Vmatrix}, \quad f = \begin{Vmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{Vmatrix}.$$

В общем случае $a_{ij} = m_{ij}p^2 + n_{ij}p + l_{ij}$, $p = \frac{d}{dt}$. При этом уравнение движения системы с инерционными элементами будет таким:

$$(Mp^2 + Np + L)\,x = f,$$

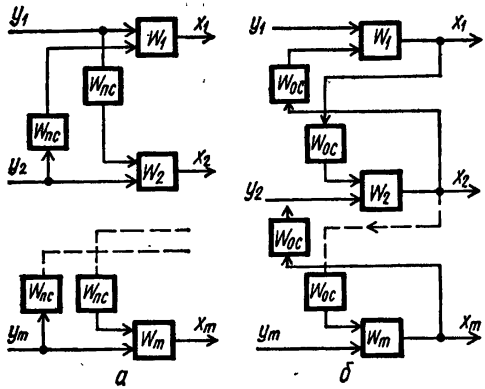

Рис. 86

где M, N, L — числовые матрицы.

Для того чтобы получить характеристический полином, нужно записать в уравнении вместо $x = Wy$ значение $Лy$, где $Л$ — скалярная матрица.

Тогда матрица $\|ЛE - W\|$ называется характеристической, а ее определитель $|ЛE - W| = 0$ будет давать характеристическое уравнение.

Для многомерных автоматических систем характерен структурный анализ. Среди элементов структуры выделяют многополюсники 1-го (рис. 86, *а*) и 2-го родов (рис. 86, *б*).

Зависимости между выходными и входными величинами многополюсников записываются с помощью матриц, а именно:

$$x = Uy \text{ и } x = Vy,$$

где

$$U = \begin{Vmatrix} W_1 & W_{\text{по}} & \dots & W_{\text{по}} \\ W_{\text{пс}} & W_2 & \dots & W_{\text{пс}} \\ \dots & \dots & \dots & \dots \\ W_{\text{пс}} & W_{\text{по}} & \dots & W_n \end{Vmatrix}, \quad V = \begin{Vmatrix} \frac{1}{W_1} & W_{\text{ос}} & \dots & W_{\text{ос}} \\ W_{\text{ос}} & \frac{1}{W_2} & \dots & W_{\text{ос}} \\ \dots & \dots & \dots & \dots \\ W_{\text{ос}} & W_{\text{ос}} & \dots & \frac{1}{W_n} \end{Vmatrix},$$

а x и y — столбцовые матрицы порядка n.

Сложные многомерные системы составляются из последовательно соединенных многополюсников 1-го и 2-го родов и поэтому их передаточные матрицы определяются по правилу:

$$U = U_m U_{m-1} \dots U_2 U_1$$

и

$$V = V_m^{-1} V_{m-1}^{-1} \dots V_2^{-1} V_1^{-1},$$

где V_l^{-1} — обратная для V_l матрица.

Передаточная матрица замкнутой многомерной автоматической системы получается из системы уравнений:

$$\begin{cases} x = Wy + f, \\ y = -Ex, \end{cases}$$

или из равенства $\Phi = (E + W)$,
где E — единичная матрица обратных связей.

ГЛАВА VII

ВАРИАЦИОННОЕ ИСЧИСЛЕНИЕ

§ 37. ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ

Вариационное исчисление — это раздел математики, который занимается исследованием на экстремум переменных величин, называемых функционалами. Оно нашло широкое применение при решении многих задач современной науки и техники, в том числе в автоматике и технической кибернетике.

Вариационное исчисление возникло на рубеже XVII и XVIII веков. Решающее влияние на развитие этой математической дисциплины оказали следующие три задачи:

а) Задача о брахистохроне *(линии наибыстрейшего ската)*. Среди всех линий, соединяющих две точки M_1 и M_2, требуется найти ту, по которой материальная точка, двигаясь под влиянием силы тяжести из M_1 без начальной скорости, достигает

M_2 в кратчайшее время. Предполагается, что точки M_1 и M_2 не лежат на одной вертикали и точка M_2 расположена ниже, чем M_1.

б) Задача о геодезических линиях. Среди всех линий, соединяющих точки M_1 и M_2 и лежащих на заданной поверхности, найти линию наименьшей длины (такие линии называются *геодезическими*).

в) Изопериметрическая задача. На плоскости найти замкнутую линию заданной длины, ограничивающую наибольшую площадь.

Эти три задачи были решены в конце XVII века разными математиками. Но общие методы решения подобных задач разработал Л. Эйлер, и его по праву считают создателем вариационного исчисления.

Рассмотрим сначала задачу о брахистохроне. Для решения этой задачи необходимо рассмотреть все возможные кривые, соединяющие точки M_1 и M_2, и из этих кривых выбрать ту, которая требуется по условию задачи.

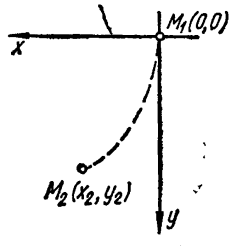

Рис. 87

Задаче можно придать аналитическую форму. Для этого надо прежде всего выбрать систему координат. Выбор этот совершенно произвольный, но для упрощения рассуждений и выкладок возьмем декартову прямоугольную систему координат, начало координат поместим в точке M_1, ось абсцисс расположим горизонтально, выбрав положительное направление влево, ось ординат расположим вертикально, выбрав положительное направление вниз (рис. 87).

Пусть точка M_2 имеет координаты x_2 и y_2. Тогда точки M_1 (0, 0) и M_2 (x_2, y_2) определяет концы всех линий, подлежащих изучению. Пусть далее искомая линия l имеет уравнение $y = y(x)$, причем функция $y(x)$ непрерывно дифференцируема на отрезке $[0; x_2]$.

Очевидно, на концах отрезка эта функция должна удовлетворять граничным условиям $y(0) = 0$, $y(x_2) = y_2$.

Из механики известно, что в любом месте линии l скорость перемещения материальной точки, отпущенной без начальной скорости, определяется по формуле

$$\frac{1}{2}v^2 = gy, \text{ или } v = \sqrt{2gy}, \tag{1}$$

где v — скорость перемещения точки, g — ускорение силы тяжести, y — путь по вертикали, пройденный точкой с момента начала движения.

Время dT, необходимое для прохождения материальной точкой участка dl вдоль кривой l, будет равно

$$dT = \frac{dl}{v} = \frac{\sqrt{1+y'^2}dx}{\sqrt{2gy}}.$$

Полное время T ската точки из M_1 в M_2 определяется интегралом:

$$T = \int_0^{x_2} \frac{dl}{v} = \frac{1}{\sqrt{2g}} \int_0^{x_2} \sqrt{\frac{1+y'^2}{y}}\, dx.$$

Итак,

$$T = \frac{1}{\sqrt{2g}} \int_0^{x_2} \sqrt{\frac{1+y'^2}{y}}\, dx \qquad (2)$$

при условии $\begin{cases} y(0) = 0, \\ y(x_2) = y_2. \end{cases}$

Так как интегрирование ведется по x, то значение T будет зависеть от выбора функции $y = y(x)$. Здесь необходимо выбрать функцию $y = y(x)$ такой, чтобы значение T было наименьшим.

Аналогичное заключение получаем и для задачи о геодезических линиях.

Если $\varphi(x, y, z) = 0$ — уравнение заданной поверхности, уравнения искомой линии наименьшей длины взяты в параметрической форме

$$\begin{cases} x = x(t), \\ y = y(t), \\ z = z(t) \end{cases} \left(dl = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\, dt \right),$$

и точки $M_1(x_1, y_1, z_1)$ и $M_2(x_2, y_2, z_2)$ соответствуют значениям t_1 и t_2 параметра t, то надо исследовать на минимум интеграл $l = \int_{t_1}^{t_2} \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}\, dt$ при условии, что $\varphi(x, y, z) = 0$.

Такой же подход будет и при приведении к аналитической форме изопериметрической задачи.

Легко заметить, что во всех трех задачах исследуемыми на экстремум величинами являются не функции в обычном понимании с точки зрения математического анализа, а какие-то математические величины иной природы, зависящие от выбора той или иной функции.

Подчеркиваем, что здесь имеет место зависимость исследуемой величины не от значения функции, а от самой выбранной функции. Такие величины и называются ф у н к ц и о н а л а м и. Функционалами являются: площадь $S = \int_a^b y dx$ криволинейной трапеции, ограниченной линиями $x = a = \text{const}$, $x = b = \text{const}$, $y = 0$ и $y = y(x)$; объем $V = \iint_D z(x, y)\, dxdy$, зависящий от выбора поверхности $z = z(x, y)$; координаты центра тяжести некоторой кривой, поверхности, тела; моменты инерции; сопротивление среды движущемуся телу в зависимости от формы последнего и т. д.

Независимо от своего содержания все функционалы воплощают в себе конкретные случаи следующей общей картины.

Пусть M — множество каких угодно объектов s (природа их безразлична; это могут быть точки, линии, поверхности, функции, совокупности функций и т. д.). Эти объекты s будем называть элементами множества M.

Определение. *Если каждому элементу s из множества M соответствует некоторое числовое значение переменной величины v из множества значений V, то на множестве M определен функционал $v[s]$.*

В зависимости от особенностей объектов s и характера зависимости значений v от выбора элементов s получается тот или иной смысл функционала, его природа.

Основной задачей вариационного исчисления является исследование на экстремум функционалов, построенных на множестве функций (совокупностей функций) из множества M. Решение такой задачи имеет сходство с решением задач на экстремум в математическом анализе. Однако существенные различия между функциями в математическом анализе и функционалами требуют и различного подхода. Этим объясняется возникновение вариационного исчисления.

В роли аргументов функционалов всегда выступают функции (их иногда называют функциональными аргументами).

Поскольку в вариационном исчислении интерес представляют только те функции, на которых функционал может иметь экстремальные значения, то из всего бесконечного многообразия функций, на которых существует функционал, выбирается лишь некоторая часть — класс допустимых функций. Выделение такого класса достигается каким-либо дополнительным условием (например, прохождение графиков через какую-либо точку, огибание заданной области и т. д.). Остальные функции из рассмотрения исключаются.

Рассмотрим подробнее функционалы $v(y(x))$, зависящие от выбора одной функции одной независимой переменной.

Так как поиск экстремума функционала осуществляется перебором, то в классе допустимых функций выбирают кривую $y(x)$ и близкие к ней кривые сравнения $\bar{y}(x)$ (эти кривые называют также смежными кривыми).

Разность $y(x) - \bar{y}(x)$ принято обозначать $\delta y(x)$ (или δy) и называть вариацией функции $y(x)$. В общих точках функций $y(x)$ и $\bar{y}(x)$ вариация δy обращается в нуль.

Термин «близкие кривые» понимается по-разному. Различают близость нулевого, первого, второго, ..., k-го порядков. Две кривые $y = y(x)$ и $y = \bar{y}(x)$ имеют близость k-го порядка, если одновременно выполняются условия:

$$|y(x) - \bar{y}(x)| < \varepsilon_0,$$

$$|y'(x) - \bar{y}'(x)| < \varepsilon_1,$$

$$|y''(x) - \bar{y}''(x)| < \varepsilon_2,$$

. . . . . .

$$|y^{(k)}(x) - \bar{y}^{(k)}(x)| < \varepsilon_k,$$

где ε_0, ε_1, ε_2, ..., ε_k — произвольно малые положительные числа и штрихами обозначены производные.

Чтобы подчеркнуть малость вариации, обычно записывают вместо δy выражение $\alpha\delta y$, где α — малый параметр.

Экстремум функционала отыскивается на множестве близких допустимых функций.

О п р е д е л е н и е. *Если экстремум функционала доставляют допустимые функции нулевого порядка близости, то экстремум называется с и л ь н ы м; если экстремум достигается на множестве допустимых функций первого порядка близости, то его называют с л а б ы м.*

Различают также абсолютный и относительный экстремум. Если кривая $y(x)$ доставляет экстремум функционалу $v(y(x))$ по сравнению с его значениями на всех допустимых кривых данного класса, то экстремум называется а б с о л ю т н ы м. Экстремум функционала $v(y(x))$ только лишь над классом близких функций называется о т н о с и т е л ь н ы м.

Из приведенных определений следует, что сильный экстремум одновременно является и слабым и что абсолютный экстремум одновременно является относительным. Относительный экстремум может оказаться как сильным, так и слабым.

Ниже рассматриваются только относительные экстремумы.

Для исследования функции на экстремум в математическом анализе используются понятия непрерывной функции, линейной функции и дифференциалов функции.

Аналогичные понятия вводятся и в вариационном исчислении для исследования на экстремум функционалов.

В математическом анализе функция $y = y(x)$ называется н е п р е р ы в н о й, если бесконечно малому приращению Δx аргумента соответствует бесконечно малое приращение $\Delta y(x)$ функции.

В вариационном исчислении функционал $v(y, x))$ называется н е п р е р ы в н ы м, если бесконечно малой вариации $\alpha\delta y(x)$ функции $y(x)$ соответствует бесконечно малое приращение функционала

$$\text{при} \quad \left.\begin{aligned} |\Delta v| &= |v(y_2(x)) - v(y_1(x))| < \varepsilon \\ |\alpha\delta y| &= |y_2(x) - y_1(x)| < \eta, \end{aligned}\right\} \tag{3}$$

где ε — сколь угодно малое произвольно заданное положительное число, а выбор η зависит от принятого значения ε.

Непрерывность функционала рассматривается также и в смысле близости k-го порядка.

Понятие линейности функционала сходно с понятием линейности функции в математическом анализе, а именно: функционал

$L(y(x))$ называется л и н е й н ы м, если он удовлетворяет следующим двум условиям:

$$\left.\begin{aligned} &1)\ L(Cy(x)) = CL(y(x)) \quad (C = \text{const}), \\ &2)\ L(y_1(x) + y_2(x)) = L(y_1(x)) + L(y_2(x)). \end{aligned}\right\} \tag{4}$$

Например, функционал

$$L(y(x)) = \int_{x_1}^{x_2} y(x)\,dx$$

линейный, так как

$$L(Cy(x)) = \int_{x_1}^{x_2} Cy(x)\,dx = C\int_{x_1}^{x_2} y(x)\,dx = CL(y(x)) \quad (C = \text{const})$$

и

$$L(y_1(x) + y_2(x)) = \int_{x_1}^{x_2} (y_1(x) + y_2(x))\,dx = \int_{x_1}^{x_2} y_1(x)\,dx + \int_{x_1}^{x_2} y_2(x)\,dx = L(y_1(x)) + L(y_2(x)).$$

Дифференциалу функции в вариационном исчислении подобно понятие вариации функционала.

Дифференциалам различных порядков, определенным для функций в математическом анализе, в вариационном исчислении подобны понятия вариаций различных порядков.

О п р е д е л е н и е. *Если при малой вариации $\alpha\delta y$ аргумента приращение Δv функционала $v(y(x))$ может быть представлено в виде суммы главной части — функционала, линейного относительно вариации аргумента,— и бесконечно малых высших порядков, т. е.*

$$\Delta v = v(y(x) + \alpha\delta y) - v(y(x)) = L(y(x);\ \alpha\delta y) + \beta(y;\ \alpha\delta y)\max|\alpha\delta y| \tag{5}$$

(причем $\beta \to 0$ при $\max|\alpha\delta y| \to 0$, α — малый параметр), то функционал $v(y(x))$ называется в а р ь и р у е м ы м ф у н к ц и о н а л о м, а главная часть его приращения, линейная относительно вариации $\alpha\delta y$ аргумента, называется п е р в о й в а р и а ц и е й ф у н к ц и о н а л а $v(y(x))$ и обозначается δv.

Таким образом,

$$\delta v(y(x)) = L(y(x);\ \alpha\delta y) = \alpha L(y;\ \delta y) \tag{6}$$

в силу линейности функционала L. Так как

$$\frac{\partial}{\partial\alpha}(v(y(x) + \alpha\delta y))|_{\alpha=0} = \lim_{\Delta\alpha\to 0}\left(\frac{\Delta v}{\Delta\alpha}\right)_{\alpha=0} = \lim_{\alpha\to 0}\frac{\Delta v}{\alpha} =$$

$$= \lim_{\alpha\to 0}\frac{L(y;\ \alpha\delta y) + \beta(y;\ \alpha\delta y)\max|\alpha\delta y|}{\alpha} =$$

$$= \lim_{\alpha\to 0}\frac{\alpha L(y;\ \delta y) + \beta(y;\ \alpha\delta y)\,|\alpha|\max|\delta y|}{\alpha} = L(y;\ \delta y),$$

то

$$\delta v(y(x)) = \alpha \cdot \frac{\partial v(y(x) + \alpha\delta y)}{\partial \alpha}\Big|_{\alpha=0}. \quad (7)$$

Если зафиксировать величины $y(x)$ и $\delta y(x)$, входящие в состав функционала, то функционал станет функцией одной переменной — параметра α:

$$v(y + \alpha\delta y) = \Phi(\alpha).$$

Допустим, что существуют производные этой функции всех порядков по α и что эту функцию можно разложить в ряд Маклорена по степеням параметра α, тогда

$$v(y(x) + \alpha\delta y) = v(y(x)) + \frac{\alpha}{1!}\left(\frac{d\Phi}{d\alpha}\right)_{\alpha=0} + \frac{\alpha^2}{2!}\left(\frac{d^2\Phi}{d\alpha^2}\right)_{\alpha=0} + \frac{\alpha^3}{3!}\left(\frac{d^3\Phi}{d\alpha^3}\right)_{\alpha=0} + \dots, \quad (8)$$

Отсюда

$$\Delta v = \frac{\alpha}{1!}\left(\frac{\partial v}{\partial \alpha}\right)_{\alpha=0} + \frac{\alpha^2}{2!}\left(\frac{\partial^2 v}{\partial \alpha^2}\right)_{\alpha=0} + \alpha^3 r_2(\alpha), \quad (9)$$

где $\alpha^3 r_2(\alpha)$ — остаток ряда.

Здесь, как и прежде, $\alpha\left(\frac{\partial v}{\partial \alpha}\right)_{\alpha=0} = \delta v(y(x))$ — первая вариация функционала.

Величина $\alpha^2\left(\frac{\partial^2 v}{\partial \alpha^2}\right)_{\alpha=0}$ называется второй вариацией функционала и обозначается $\delta^2 v(y(x))$:

$$\delta^2 v(y(x)) = \alpha^2\left(\frac{\partial^2 v}{\partial \alpha^2}\right)_{\alpha=0}. \quad (10)$$

Точно так же можно ввести вариации третьего и более высоких порядков.

Итак, разложение функционала в ряд Маклорена по параметру α дает:

$$\Delta v(y(x)) = \delta v + \frac{1}{2!}\delta^2 v + \alpha^3 r_2(\alpha). \quad (11)$$

§ 38. ЭКСТРЕМУМ ФУНКЦИОНАЛА. НЕОБХОДИМОЕ УСЛОВИЕ ЭКСТРЕМУМА

Определение. *Функционал $v(y(x))$ достигает на кривой $y = y_0(x)$ минимума, если значения функционала $v(y(x))$ на любой кривой, близкой к кривой $y = y_0(x)$ не меньше, чем $v(y_0(x))$, т. е. $\Delta v = v(y(x)) - v(y_0(x)) \geqslant 0$ (если $\Delta v = 0$ только при $y(x) = y_0(x)$, то на этой кривой достигается строгий минимум).*

Аналогично определяется максимум функционала.

Так же, как и в математическом анализе, понятия минимума и максимума объединяются термином экстремум функционала.

В связи с тем что выводы, касающиеся минимума и максимума функционала, сходны между собой и отличаются в формулах только знаками, нет необходимости проводить специальное исследование каждого из двух видов экстремумов. Так как вариационные задачи связаны преимущественно с минимальными значениями, то обычно ограничиваются изучением минимума функционала. Этой традиции далее будем следовать и мы.

В математическом анализе изучаются необходимые и достаточные условия, определяющие экстремум функции. Аналогичное исследование экстремума функционала осуществляется и в вариационном исчислении. Однако в отличие от математического анализа здесь можно в большинстве случаев, исходя из конкретных физических условий задачи, установить (как только выполнено необходимое условие существования экстремума), имеет ли место экстремум, и если да, то и его характер. Поэтому в вариационном исчислении ведущую роль играет изучение необходимого условия существования экстремума.

Такому положению способствует и сложность достаточных условий, которые применяются, как правило, лишь в тех случаях, когда необходимо глубокое и тонкое исследование экстремума.

Теорема о необходимом условии экстремума функционала.

Теорема. *Если функционал* $v\,(y\,(x))$ *имеет вариацию и достигает экстремума на кривой* $y = y_0\,(x)$, *то вариация функционала на этой кривой равна нулю, т. е.*

$$\delta v\,(y_0\,(x)) = 0. \tag{12}$$

Доказательство. Если функция $y\,(x)$ и ее вариация фиксированы (а это и имеет место для кривой $y = y_0\,(x)$), то функционал на близкой кривой $y = y_0\,(x) + \alpha\delta y$ имеет вид $v = v\,(y_0\,(x) + \alpha\delta y)$, т. е. превращается в функцию параметра α. Необходимым условием существования экстремума такой функции $\Phi\,(\alpha) = v\,(y_0\,(x) + \alpha\delta y)$, как известно, является равенство нулю ее первой производной на кривой $y = y_0\,(x)$

$$\frac{d\Phi}{d\alpha}\Big|_{\alpha=0} = \frac{\partial v\,(y_0\,(x) + \alpha\delta y)}{\partial\alpha}\Big|_{\alpha=0} = 0.$$

Так как

$$\alpha\left(\frac{\partial}{\partial\alpha}\, v\,(y_0\,(x) + \alpha\delta y)\,|_{\alpha=0}\right) = \delta v\,(y_0\,(x)),$$

то

$$\delta v\,(y_0\,(x)) = 0.$$

Определение. *Кривые, на которых выполнено необходимое условие экстремума, называются экстремалями функционала.*

По своему содержанию термин «экстремаль» подобен термину «точка стационарности» в математическом анализе.

Замечание. Полученное необходимое условие существования экстремума почти без всяких изменений распространяется на функционалы любых видов и от любого количества независимых функций.

Простейшая вариационная задача. Уравнение Эйлера. Среди всех функционалов простейшим является функционал вида

$$v(y(x)) = \int_{x_1}^{x_2} F(x, y(x), y'(x))\,dx. \tag{13}$$

Отыскание экстремалей (кривых, на которых может реализоваться экстремум) такого функционала составляет простейшую вариационную задачу.

В дальнейшем предполагается, что подынтегральная функция $F(x, y(x), y'(x))$ трижды дифференцируема по всем своим аргументам; допустимые функции закреплены на концах интервала интегрирования ($y(x_1) = y_1$, $y(x_2) = y_2$) и непрерывно дифференцируемы внутри этого интервала.

Основная лемма вариационного исчисления. *Если*

$$\int_{x_1}^{x_2} M(x)\,\eta(x)\,dx = 0, \tag{14}$$

причем $M(x)$ — фиксированная функция, непрерывная в интервале интегрирования $[x_1; x_2]$, и функция $\eta(x)$, непрерывная в этом интервале вместе со своей первой производной, равна нулю на концах интервала ($\eta(x_1) = \eta(x_2) = 0$), а в остальном совершенно произвольна, то функция $M(x)$ тождественно равна нулю в интервале $[x_1; x_2]$.

Доказательство. Предположим, что утверждение леммы неверно и что в некоторой точке ξ интервала $[x_1; x_2]$ функция $M(\xi) \neq 0$, например, $M(\xi) > 0$. Тогда вследствие непрерывности она будет положительна и в некоторой окрестности $[\xi_1; \xi_2]$ точки ξ, причем

$$x_1 < \xi_1 < \xi < \xi_2 < x_2.$$

Выбор функции $\eta(x)$, в основном, произволен. Поэтому определим ее так:

$$\eta(x) = \begin{cases} 0 & \text{при } x_1 \leqslant x \leqslant \xi_1, \\ \varphi^2(x) & \text{при } \xi_1 \leqslant x \leqslant \xi_2, \\ 0 & \text{при } \xi_2 \leqslant x \leqslant x_2, \end{cases}$$

где $\varphi(x)$ — произвольная действительная и непрерывная вместе со своей первой производной функция, причем $\varphi(\xi_1) = \varphi(\xi_2) = 0$, но $\varphi(x) \neq 0$, если $x \in]\xi_1; \xi_2[$.

Функция $\eta(x)$, выбранная таким образом, удовлетворяет требованиям леммы.

Получаем

$$\int_{x_1}^{x_2} M(x)\,\eta(x)\,dx = \int_{\xi_1}^{\xi_2} M(x)\,\varphi^2(x)\,dx.$$

Так как в интервале $]\xi_1; \xi_2[$ подынтегральная функция положительна ($\varphi^2(x) > 0$, и по предположению $M(\xi)$ положительна), то интеграл также положителен, а по условию леммы он должен быть равен нулю. Противоречие доказывает справедливость леммы.

Заметим, что аналогичную лемму можно доказать и для кратных интегралов.

Перейдем теперь к изучению необходимого условия экстремума простейшего функционала (13):

$$v(y(x)) = \int_{x_1}^{x_2} F(x, y, y')\, dx.$$

Предположим, что экстремум функционала достигается на кривой $y = y(x)$. Для вывода необходимого условия экстремума возьмем из числа допустимых функций вспомогательную функцию вида

$$\bar{y}(x) = y(x) + \alpha\eta(x), \tag{15}$$

где α — произвольный параметр.

Так как $\bar{y}(x)$ — допустимая функция, то она непрерывно дифференцируема внутри интервала $[x_1; x_2]$, и $\bar{y}(x_1) = y_1$, $\bar{y}(x_2) = y_2$. Следовательно, функция $\eta(x)$ непрерывно дифференцируема внутри $[x_1; x_2]$ и на концах интервала принимает нулевые значения ($\eta(x_1) = \eta(x_2) = 0$). В остальном функция $\eta(x)$ совершенно произвольна.

Нетрудно установить, что функционал $v(y(x))$, вычисленный для функции $\bar{y}(x)$, превращается в функцию параметра α (обозначим ее через $\Phi(\alpha)$):

$$v(\bar{y}(x)) = \int_{x_1}^{x_2} F(x, y + \alpha\eta, y' + \alpha\eta')\, dx = \Phi(\alpha). \tag{16}$$

Так как на кривой $y(x) = \bar{y}(x)|_{\alpha=0}$ достигается экстремум функционала, то на этой кривой выполнено необходимое условие экстремума функции $\Phi(\alpha)$:

$$\frac{\partial v(\bar{y}(x))}{\partial \alpha}\bigg|_{\alpha=0} = \Phi'(\alpha)|_{\alpha=0} = \Phi(0) = 0.$$

Выполняя дифференцирование по параметру, имеем

$$\frac{\partial v(\bar{y}(x))}{\partial \alpha}\bigg|_{\alpha=0} \frac{\partial}{\partial \alpha}\left(\int_{x_1}^{x_2} F(x, y(x) + \alpha\eta(x), y'(x) + \alpha\eta'(x))\, dx\right)_{\alpha=0} =$$

$$= \left(\int_{x_1}^{x_2} \frac{\partial}{\partial \alpha} F(x, y + \alpha\eta, y' + \alpha\eta')\, dx\right)_{\alpha=0} = \int_{x_1}^{x_2} \left(\frac{\partial F}{\partial (y + \alpha\eta)}\eta +\right.$$

$$\left. + \frac{\partial F}{\partial (y' + \alpha\eta')}\eta'\right)_{\alpha=0} dx = \int_{x_1}^{x_2} \left(\frac{\partial F}{\partial y}\eta + \frac{\partial F}{\partial y'}\eta'\right) dx.$$

Поэтому

$$\int_{x_1}^{x_2}\left(\frac{\partial F}{\partial y}\eta+\frac{\partial F}{\partial y'}\eta'\right)dx=0. \tag{17}$$

Применим ко второму слагаемому уравнения (17) интегрирование по частям:

$$\int_{x_1}^{x_2}\frac{\partial F}{\partial y'}\eta'dx=\left|\begin{matrix}u=\dfrac{\partial F}{\partial y'}, & du=\dfrac{d}{dx}\left(\dfrac{\partial F}{\partial y'}\right)dx\\ dv=\eta'dx, & v=\eta\end{matrix}\right|=$$

$$=\frac{\partial F}{\partial y'}\eta\Bigg|_{x_1}^{x_2}-\int_{x_1}^{x_2}\eta\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)dx=\int_{x_1}^{x_2}\eta\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)dx,$$

так как $\eta(x_1)=\eta(x_2)=0.$

Подставляя найденный результат в необходимое условие экстремума функционала (17), получим

$$\int_{x_1}^{x_2}\left(F_y\eta-\eta\frac{d}{dx}F_{y'}\right)dx=0 \text{ или } \int_{x_1}^{x_2}\left(F_y-\frac{d}{dy}F_{y'}\right)\eta dx=0$$

(здесь и в дальнейшем $F_x=\dfrac{\partial F}{\partial x}$, $F_y=\dfrac{\partial F}{\partial y}$, $F_{y'}=\dfrac{\partial F}{\partial y'}$ и т. п.).

Отсюда (благодаря условиям, наложенным на функции $F(x, y, y')$ и $\eta(x)$) в силу основной леммы вариационного исчисления следует, что

$$F_y-\frac{d}{dx}F_{y'}=0. \tag{18}$$

Это соотношение является практически удобной формой необходимого условия экстремума простейшего функционала. Оно было впервые (в 1744 г.) найдено Л. Эйлером и называется уравнением Эйлера.

Функции $y=y(x)$, представляющие собой решения уравнения Эйлера,— это экстремали простейшего функционала.

Если учесть, что

$$\frac{d}{dx}F_{y'}(x, y, y')=F_{xy'}+F_{yy'}y'+F_{y'y'}y'',$$

то легко получить уравнение Эйлера в развернутой форме:

$$F_y-F_{xy'}-y'F_{yy'}-y''F_{y'y'}=0. \tag{19}$$

В общем случае уравнение Эйлера является нелинейным дифференциальным уравнением второго порядка. Его общее решение содержит две произвольные постоянные и представляет собой двупараметрическое семейство кривых, на котором фактически может реализоваться экстремум функционала. Из этого множества кривых надо выбрать одну кривую, проходящую через две заданные точки

(x_1, y_1) и (x_2, y_2) и удовлетворяющую граничным условиям (условиям «закрепления на концах»). Достигается это определением частных значений произвольных постоянных C_1 и C_2 в составе общего решения уравнения Эйлера.

Характер возможного экстремума определяется законом приращения функционала на выбранной экстремали. Из разложения функционала в ряд Маклорена следует, что на экстремалях (где $\delta v = 0$) знак приращения Δv определяется знаком второй вариации $\delta^2 v$ (при условии $\delta^2 v \neq 0$). Если $\delta^2 v > 0$, то возможен минимум, а если $\delta^2 v < 0$,— максимум функционала.

Можно показать (и это сделано в § 41), что если $\delta^2 v > 0$, то $F_{y'y'} \geqslant 0$, и что при $\delta^2 v < 0$ будет $F_{y'y'} \leqslant 0$. Поэтому, если на экстремалях выполняется условие $F_{y'y'} \geqslant 0$, то возможен минимум функционала, если $F_{y'y'} \leqslant 0$, то возможен максимум функционала. Это условие называется условием Лежандра. Его выполнения недостаточно для существования экстремума функционала.

Некоторые частные случаи интегрируемости уравнений Эйлера. В общем случае уравнение Эйлера (18) не интегрируется в квадратурах. Поэтому особый интерес представляют те случаи, когда уравнение Эйлера интегрируется непосредственно или допускает понижение порядка. Рассмотрим некоторые из этих случаев:

1. Подынтегральная функция F не зависит от y':

$$F = F(x, y).$$

В этом случае, очевидно,

$$F_{xy'} = F_{yy'} = F_{y'y'} = 0, \quad \frac{d}{dx} F_{y'} = 0,$$

а уравнение Эйлера имеет вид $F_y = 0$. Это конечное, а не дифференциальное уравнение.

Экстремум функционала может реализоваться на кривой, определяемой этим уравнением лишь в очень редких случаях — когда кривая проходит через граничные точки (x_1, y_1) и (x_2, y_2).

Пример. Найти экстремали функционала

$$v(y(x)) = \int_{x_1}^{x_2} (y^3 + x^2 + 3)\, dx$$

при $y(x_1) = y_1$, $y(x_2) = y_2$.

Решение. Здесь $F = y^3 + x^2 + 3$ не зависит от y'.

Уравнение Эйлера имеет вид: $3y^2 = 0$ или $y = 0$.

Экстремалью может быть только прямая $y = 0$ и то лишь при условии, что $y(x_1) = y(x_2) = 0$.

В противном случае эта прямая не будет экстремалью и заданный функционал не достигает экстремума на непрерывных функциях.

2. Подынтегральная функция F линейно зависит от y' (или $F_{y'y'} \equiv 0$):

$$F(x, y, y') = P(x, y) + Q(x, y)\, y'.$$

Здесь функционал $v(y(x))$ имеет вид

$$v(y(x)) = \int_{x_1}^{x_2} (P(x, y) + Q(x, y)\, y')\, dx.$$

В этом случае уравнение Эйлера $F_y - \frac{d}{dx} F_y = 0$ запишется так:

$$\frac{\partial P(x, y)}{\partial y} + \frac{\partial Q(x, y)}{\partial y} y' - \frac{d}{dx} Q(x, y) = 0,$$

или

$$\frac{\partial P}{\partial y} + \frac{\partial Q}{\partial y} y' - \frac{\partial Q}{\partial x} - \frac{\partial Q}{\partial y} y' = 0,$$

т. е.

$$\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial x} = 0.$$

Как и в предыдущем случае, получили не дифференциальное, а конечное уравнение. Кривая, определяемая этим уравнением, вообще говоря, не удовлетворяет граничным условиям. Поэтому в общем случае здесь вариационная задача не имеет решения в классе непрерывных функций.

Если же $\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial x} \equiv 0$, то выражение $Pdx + Qdy$ является полным дифференциалом. Тогда интеграл не зависит от формы пути интегрирования, а функционал

$$v(y(x)) = \int_{x_1}^{x_2} \left(P(x, y) + Q(x, y) \frac{dy}{dx}\right) dx = \int_{x_1}^{x_2} (Pdx + Qdy)$$

не зависит от выбора кривой $y = y(x)$, и вариационная задача теряет смысл (теряет смысл сама задача отыскания экстремума функционала, имеющего постоянное значение).

Таким образом, при $F_{y'y'} \equiv 0$ всегда имеется вырожденный функционал.

Замечание. В общем случае, когда $F_{y'y'} \not\equiv 0$ вследствие непрерывности частных производных от функции F, всегда будет непрерывна и производная $y'(x)$, и экстремали не будут иметь угловых точек. Если $F_{y'y'}$ в отдельных точках обращается в нуль (но $F_{y'y'} \not\equiv 0$), то возможны изломы экстремалей.

Примеры

1. Найти экстремали функционала

$$v(y(x)) = \int_{-1}^{0} (xy - xy^2y')\, dx; \quad y(-1) = 1,\ y(0) = a.$$

Р е ш е н и е. Здесь $F = xy - xy^2y'$ зависит от y' линейно. Уравнение Эйлера $\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial x} = 0$ при $P(x, y) = xy$ и $Q(x, y) = -xy^2$, принимает вид: $x + y^2 = 0$, или $y^2 = -x$.

Первое граничное условие $y(-1) = 1$ удовлетворяется, второе — только при $a = 0$. Если $a \neq 0$, то экстремали, удовлетворяющие граничным условиям, не существуют.

2. Найти экстремали функционала

$$v(y(x)) = \int_0^1 (x^2y + y^2y')\,dx; \quad y(0) = 0, \; y(1) = 1.$$

Р е ш е н и е. Здесь уравнение Эйлера имеет вид $x^2 = 0$, или $x = 0$.

Первое граничное условие удовлетворяется, так как уравнение экстремали вообще не зависит от y. Второе граничное условие не выполняется никогда. Следовательно, экстремали, удовлетворяющие заданным граничным условиям, не существуют.

3. Найти экстремали функционала

$$v(y(x)) = \int_{x_1}^{x_2} (xy^2 + x^2yy')\,dx; \quad y(x_1) = y_1, \; y(x_2) = y_2.$$

Р е ш е н и е. Здесь $\frac{\partial P}{\partial y} = 2xy$, $\frac{\partial Q}{\partial x} = 2xy$. Отсюда следует, что $\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial x} \equiv 0$, и под знаком интеграла стоит полный дифференциал:

$$v(y(x)) = \int_{x_1}^{x_2} (xy^2 + x^2yy')\,dx = \frac{1}{2}\int_{x_1}^{x_2} d(x^2y^2) = \frac{1}{2}(x_2^2y_2^2 - y_1^2y_1^2).$$

Интеграл не зависит от формы пути интегрирования, а зависит только от граничных условий. Вариационная задача не имеет смысла.

3. Подынтегральная функция F зависит непосредственно только от y' : $F = F(y')$.

В этом случае $F_y = F_{xy'} = F_{yy'} = 0$, и поэтому уравнение Эйлера будет иметь вид

$$y''F_{y'y'} = 0.$$

Это уравнение распадается на два уравнения: $y'' = 0$ и $F_{y'y'} = 0$.

Если $y'' = 0$, то $y = C_1x + C_2$. Имеем двупараметрическое семейство прямых.

Если уравнение $F_{y'y'}(y') = 0$ имеет действительные корни $y' = k$, то $y = kx + C$. Получаем однопараметрическое семейство прямых, входящее в найденное выше двупараметрическое семейство.

Если же уравнение $F_{y'y'}(y') = 0$ имеет комплексные корни $y' = \alpha + \beta i$, то $y = (\alpha + \beta i)x + C$ не может представлять собой экстремали, так как все исследование осуществляется в области только вещественных переменных.

Пример. Найти экстремали функционала

$$v(y(x)) = \int_{x_1}^{x_2} \sqrt{1 + y^2}\,dx; \quad y(x_1) = y_1, \; y(x_2) = y_2.$$

Р е ш е н и е. Здесь $F = \sqrt{1 + y'^2}$ зависит непосредственно только от y'. Уравнение Эйлера $y''F_{y'y'} = 0$ будет иметь вид

$$\frac{y''}{(\sqrt{1 + y'^2})} = 0, \text{ или } y'' = 0.$$

Отсюда $y = C_1x + C_2$. Используя граничные условия, находим частные значения C_1 и C_2:

$$C_1 = \frac{y_2 - y_1}{x_2 - x_1}, \quad C_2 = \frac{y_1x_2 - y_2x_1}{x_2 - x_1}.$$

Поэтому

$$y = \frac{y_2 - y_1}{x_2 - x_1} x + \frac{y_1x_2 - y_2x_1}{x_2 - x_1},$$

или

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1}.$$

Следовательно, существует одна экстремаль — прямая, проходящая через граничные точки.

Так как здесь $F_{y'y'} = \frac{1}{(1 + y'^2)^{3/2}} > 0$, то по условию Лежандра на найденной экстремали может реализоваться минимум функционала.

Заметим, что записанный в этом примере функционал выражает длину дуги, заключенной между двумя граничными точками. Из полученного решения следует, что наименьшее расстояние между двумя точками (если нет дополнительных условий) может достигаться только на прямой — факт, принятый в элементарной геометрии как аксиома.

4. Подынтегральная функция F зависит непосредственно только от x и y' : $F = F(x, y')$.

В этом случае $F_y = 0$, и уравнение Эйлера $F_y - \frac{d}{dx} F_{y'} = 0$ принимает вид

$$\frac{d}{dx} F_{y'} = 0.$$

Оно имеет первый интеграл:

$$F_{y'} = C_1. \tag{20}$$

Получившееся дифференциальное уравнение первого порядка можно проинтегрировать или непосредственно (так как y в уравнение не входит), или введением вспомогательной переменной (параметра).

Пример. Найти экстремали функционала

$$v(y(x)) = \int_{x_1}^{x_2} (xy' + y'^2)\, dx.$$

Р е ш е н и е. Здесь $F = xy' + y'^2$ не зависит непосредственно от y. Поэтому уравнение Эйлера имеет первый интеграл $F_{y'} = C_1$. Так как $F_{y'} = x + 2y'$, то $x + 2y' = C_1$.

Непосредственно интегрируя это уравнение первого порядка, находим

$$\frac{x^2}{2} + 2y = C_1 x + C_2,$$

или

$$y = -\frac{x^2}{4} + \frac{1}{2} C_1 x + \frac{C_2}{2}.$$

Полученное означает, что экстремали заданного функционала составляют двупараметрическое семейство парабол, причем, частные значения произвольных постоянных C_1 и C_2 можно подобрать так, чтобы удовлетворить любым конечным граничным условиям.

Так как здесь $F_{y'y'} = 2 > 0$, то, в соответствии с условием Лежандра, на этих экстремалях может реализоваться минимум функционала.

5. Подынтегральная функция F зависит непосредственно только от y и $y' : F(y, y')$.

В этом случае $F_{xy'} = 0$, и уравнение Эйлера (19) принимает вид

$$F_y - y' F_{yy'} - y'' F_{y'y'} = 0.$$

Умножая обе части равенства на y', получаем

$$y' F_y - y'^2 F_{yy'} - y' y'' F_{y'y'} = 0,$$

или

$$\frac{d}{dx} F(y, y') - \frac{d}{dx} (y' F_{y'}) = 0,$$

т. е.

$$\frac{d}{dx} (F - y' F_{y'}) = 0.$$

Это уравнение имеет первый интеграл вида

$$F - y' F_{y'} = C_1. \tag{21}$$

Получившееся дифференциальное уравнение первого порядка оказывается возможным проинтегрировать либо разделением переменных, либо с помощью параметра.

Пример. (*Задача о брахистохроне*). Найти экстремаль функционала

$$v(y(x)) = \frac{1}{\sqrt{2g}} \int_0^{x_2} \sqrt{\frac{1 + y'^2}{y}}\, dx,$$

удовлетворяющую граничным условиям $y(0) = 0$, $y(x_2) = y_2$.

Решение. Так как $F = \sqrt{\frac{1 + y'^2}{y}}$ зависит только от y и y', то можно сразу записать первый интеграл уравнения Эйлера $F - y' F_{y'} = C_1$.

Имеем

$$F_{y'} = \frac{y'}{\sqrt{y(1 + y'^2)}}.$$

Поэтому первый интеграл уравнения Эйлера для рассматриваемого функционала будет

$$\sqrt{\frac{1 + y'^2}{y}} - \frac{y'^2}{\sqrt{y(1 + y'^2)}} = \overline{C}_1,$$

или, после упрощения,

$$y(1+y'^2)=C_1 \quad \left(C_1=\frac{1}{C_1^2}\right).$$

Введем параметр t, полагая $y'=\operatorname{ctg} t$. Тогда

$$\begin{cases} y(1+\operatorname{ctg}^2 t)=C_1, \\ \dfrac{dy}{dx}=\operatorname{ctg} t, \end{cases} \quad \text{или} \quad \begin{cases} y=C_1\sin^2 t, \\ dx=\dfrac{dy}{\operatorname{ctg} t}. \end{cases}$$

Так как

$$dy=2C_1\sin t\cos t\,dt,$$

то

$$dx=\frac{2C_1\sin^2 t\cos t dt}{\cos t}=2C_1\sin^2 t\,at.$$

Итак,

$$dx=C_1(1-\cos 2t)\,dt,$$

$$x=C_1\left(t-\frac{1}{2}\sin 2t\right)+C_2.$$

Поэтому

$$\begin{cases} x-C_2=\dfrac{C_1}{2}(2t-\sin 2t), \\ y=\dfrac{C_1}{2}(1-\cos 2t). \end{cases}$$

Заменим $2t=t_1$ и учтем граничное условие $y(0)=0$. Тогда найдем, что $C_2=0$, и окончательно получим

$$\begin{cases} x=\dfrac{C_1}{2}(t_1-\sin t_1) \\ y=\dfrac{C_1}{2}(1-\cos t_1). \end{cases}$$

Эта система параметрических уравнений определяет семейство циклоид, причем $\frac{C_1}{2}$ является радиусом катящегося круга.

Так как здесь $F_{y'y'}=\dfrac{1}{\sqrt{y(1+y'^2)^{3/2}}}>0$, то по условию Лежандра на полученных экстремалях возможен минимум функционала (по содержанию задачи — минимальный расход времени). Исходя из физических условий заключаем: линией наибыстрейшего ската материальной точки в положение M_2 (рис. 87), отпущенной без начальной скорости из положения M_1 (брахистохроной), будет циклоида. Ее размеры $\left(\text{радиус катящегося круга } \frac{C_1}{2}\right)$ определяются взаимным положением точек M_1 и M_2.

Функционалы других видов. Другие типы вариационных задач. До сих пор мы рассматривали лишь простейший функционал, который зависит от выбора одной функции только одной переменной.

В практике же могут встретиться и такие функционалы, которые зависят от выбора одной функции нескольких аргументов или от выбора нескольких функций одного и того же аргумента, или, наконец, от выбора нескольких функций, каждая из которых является функцией двух, трех и более аргументов, одинаковых для всех функций. Могут встретиться также функционалы, в которых, помимо первой производной, участвуют производные второго, третьего и более высоких порядков, двойные, тройные интегралы и другие случаи.

Кроме того, и сами вариационные задачи могут быть без задания фиксированных граничных условий (с подвижными границами).

Подход к таким функционалам и решение для них вариационных задач принципиально те же, что и для рассмотренного простейшего функционала и простейшей вариационной задачи.

1. Функционалы, зависящие от нескольких функций и их первых производных.

Если дан функционал

$$v = \int_{x_1}^{x_2} F(x;\ y_1, y_2, \ldots, y_n;\ y_1', y_2', \ldots, y_n')\, dx \qquad (22)$$

при граничных условиях

$$y_i(x_1) = y_{i_1},\ y_i(x_2) = y_{i_2} \quad (i = 1, 2, 3, \ldots, n), \qquad (23)$$

то, варьируя по очереди каждую из функций y_i ($i = 1, 2, 3, \ldots, n$) при фиксированных остальных, приходим к системе уравнений Эйлера:

$$F_{y_i} - \frac{d}{dx} F_{y_i'} = 0 \quad (i = 1, 2, 3, \ldots, n). \qquad (24)$$

Для этой системы дифференциальных уравнений решение (экстремаль функционала) зависит от $2n$ произвольных постоянных, значения которых определяются по граничным условиям.

Геометрическое истолкование решения получается в абстрактном $(n+1)$-мерном пространстве. Характер возможного экстремума устанавливается с помощью следующей системы неравенств — условий Лежандра: *чтобы на экстремали был возможен минимум, необходимо:*

$$F_{y_1'y_1'} \geqslant 0,\ \begin{vmatrix} F_{y_1'y_1'} & F_{y_1'y_2'} \\ F_{y_2'y_1'} & F_{y_2'y_2'} \end{vmatrix} \geqslant 0,\ \ldots,\ \begin{vmatrix} F_{y_1'y_1'} & \ldots & F_{y_1'y_n'} \\ \vdots & & \vdots \\ F_{y_n'y_1'} & \ldots & F_{y_n'y_n'} \end{vmatrix} \geqslant 0. \qquad (25)$$

Пример. Найти экстремали функционала

$$v = \int_0^{\pi} (9y'^2 + 2yz + 16z'^2 + 2yz)\, dx$$

при граничных условиях

$$y(0)=4,\ z(0)=1,\ y(\pi)=4\left(e^{\frac{\pi}{2}}-\frac{1}{9}\right),\ z(\pi)=e^{\frac{\pi}{2}}+1.$$

Р е ш е н и е. Так как подынтегральная функция

$$F=9y'^2+2y^2+16z'^2+2yz$$

зависит от двух различных функций одного и того же аргумента и их производных первого порядка, то надо составить систему двух уравнений Эйлера:

$$\begin{cases} F_y-\dfrac{d}{dx}F_{y'}=0. \\ F_z-\dfrac{d}{dx}F_{z'}=0. \end{cases}$$

Имеем:

$$F_y=4y+2z,\ F_{y'}=18y',\ \frac{d}{dx}F_{y'}=18y'',$$

$$F_z=2y,\ F_{z'}=32z',\ \frac{d}{dx}F_{z'}=32z''.$$

В результате получаем такую систему уравнений:

$$\begin{cases} 4y+2z-18y''=0, \\ 2y-32z''=0, \end{cases} \text{ или } \begin{cases} 2y+z-9y''=0, \\ y-16z''=0. \end{cases}$$

Решая эту систему, находим

$$z=C_1e^{\frac{x}{2}}+C_2e^{-\frac{x}{2}}+C_3\cos\frac{x}{6}+C_4\sin\frac{x}{6},$$

$$y=4C_1e^{\frac{x}{2}}+4C_2e^{-\frac{x}{2}}-\frac{4}{9}C_3\cos\frac{x}{6}-\frac{4}{9}C_4\sin\frac{x}{6}.$$

Используя граничные условия, получаем

$$C_1=1,\ C_2=0,\ C_3=0,\ C_4=2.$$

Поэтому искомой экстремалью будет пространственная кривая, определенная уравнениями

$$\begin{cases} y=4e^{\frac{\pi}{2}}-\dfrac{8}{9}\sin\dfrac{x}{6}, \\ z=e^{\frac{x}{2}}+2\sin\dfrac{x}{6}. \end{cases}$$

Так как

$$F_{y'y'}=18>0,\ F_{z'z'}=32>0,\ F_{y'z'}=F_{z'y'}=0,$$

то условие (25) дает

$$F_{y'y}>0,\ \begin{vmatrix} F_{y'y'} & F_{y'z'} \\ F_{z'y'} & F_{z'z'} \end{vmatrix}=\begin{vmatrix} 18 & 0 \\ 0 & 32 \end{vmatrix}=576>0;$$

следовательно, на найденной экстремали возможен минимум исследуемого функционала.

2. Функционалы, зависящие от производных второго и более высоких порядков.

Для функционалов вида

$$v = \int_{x_1}^{x_2} F(x, y(x), y'(x), y''(x), \ldots, y^{(n)}(x))\, dx \tag{26}$$

необходимое условие экстремума $\delta v = 0$ приводит к обобщению уравнения Эйлера — дифференциальному уравнению порядка $2n$:

$$F_y - \frac{d}{dx} F_{y'} + \frac{d^2}{dx^2} F_{y''} - \frac{d^3}{dx^3} F_{y'''} + \cdots + (-1)^n \frac{d^n}{dx^n} F_{y^{(n)}} = 0. \tag{27}$$

Оно известно под названием уравнения Эйлера — Пуассона. $2n$ произвольных постоянных общего решения этого уравнения определяются из граничных условий.

Пример. Найти экстремали функционала

$$v = \int_{x_1}^{x_2} (2yy' - 2y^2 - y'^2 + y''^2)\, dx.$$

Решение. Здесь $F = 2yy' - 2y^2 - y'^2 + y''^2$, поэтому

$$F_y = 2y' - 4y, \quad F_{y'} = 2y - 2y', \quad \frac{d}{dx} F_{y'} = 2y' - 2y'',$$

$$F_{y''} = 2y'', \quad \frac{d^2}{dx^2} F_{y''} = 2y^{IV}.$$

Уравнение Эйлера — Пуассона имеет вид

$$(2y' - 4y) - (2y' - 2y'') + 2y^{IV} = 0,$$

или

$$y^{IV} + y'' - 2y = 0.$$

Его решение

$$y = C_1 e^x + C_2 e^{-x} + C_3 \cos \sqrt{2}x + C_4 \sin \sqrt{2}x$$

определяет четырехпараметрическое семейство экстремалей. Частные значения произвольных постоянных вычисляются по граничным условиям.

3. Более сложные функционалы. Уравнение Эйлера — Пуассона может быть обобщено на случай функционалов, зависящих от многих функций. Если функционал v имеет вид

$$v = (y_1(x), y_2(x), \ldots, y_n(x)) =$$

$$= \int_{x_1}^{x_2} F(x;\ y_1(x), y_1'(x), \ldots, y_1^{(n_1)}(x);\ y_2(x), y_2'(x), \ldots, y_2^{(n_2)}(x);\ \ldots$$

$$\ldots;\ y_m(x), y_m'(x), \ldots, y_m^{(n_m)}(x))\, dx, \tag{28}$$

то, варьируя по очереди каждую функцию $y_i(x)$ $(i = 1, 2, 3, \ldots, m)$ при фиксированных остальных, получаем необходимое условие экстремума функционала в виде системы уравнений:

$$F_{y_i} - \frac{d}{dx} F_{y_i'} + \frac{d^2}{dx^2} F_{y_i''} + \cdots + (-1)^{n_i} \frac{d^{n_i}}{dx^{n_i}} F_{y_i^{(n_i)}} = 0 \qquad (29)$$

$$(i = 1, 2, 3, \ldots, m).$$

Если аргумент функционала зависит от нескольких переменных:

$$v(z(x_1, x_2, \ldots, x_n)) = \int_{\Omega} F(x_1, x_2, \ldots, x_n; z; p_1, p_2, \ldots$$

$$\ldots, p_n)\, dx_1 \cdot dx_2 \ldots dx_n, \qquad (30)$$

где $p_i = \frac{\partial z}{\partial x_i}$ $(i = 1, 2, 3, \ldots, n)$, а символ $\int_{\Omega}$ — это символ n-кратного интеграла по n-мерной области Ω, то необходимое условие экстремума записывается дифференциальным уравнением в частных производных:

$$F_z - \sum_{i=1}^{n} \frac{\partial}{\partial x_i} (F_{p_i}) = 0. \qquad (31)$$

Этому уравнению должна удовлетворять функция $z = z(x_1, x_2, \ldots, x_n)$, реализующая экстремум функционала v. Оно называется уравнением Остроградского.

Если подынтегральная функция функционала v зависит от производных не только первого, но и более высоких порядков, то необходимое условие экстремума функционала представляет собой дифференциальное уравнение в частных производных, подобное уравнению Эйлера — Пуассона.

Например, для функционала

$$v(z(x, y)) = \iint_D F\left(x, y, z; \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}; \frac{\partial^2 z}{\partial x^2}, \frac{\partial^2 z}{\partial x \partial y}, \frac{\partial^2 z}{\partial y^2}\right) dx dy \qquad (32)$$

необходимое условие экстремума записывается уравнением в частных производных четвертого порядка

$$F_z - \frac{\partial}{\partial x} F_p - \frac{\partial}{\partial y} F_q + \frac{\partial^2}{\partial x^2} F_r + \frac{\partial^2}{\partial x \partial y} F_s + \frac{\partial^2}{\partial y^2} F_t = 0, \qquad (33)$$

где $p = \frac{\partial z}{\partial x}$, $q = \frac{\partial z}{\partial y}$, $r = \frac{\partial^2 z}{\partial x^2}$, $s = \frac{\partial^2 z}{\partial x \partial y}$, $t = \frac{\partial^2 z}{\partial y^2}$.

§ 39. УСЛОВНЫЙ ЭКСТРЕМУМ ФУНКЦИОНАЛА

Вариационной задачей на условный экстремум называется задача, в которой требуется исследовать на экстремум функционал, когда на функции, от выбора которых этот функционал зависит,

наложены дополнительные условия, называемые связями (в частном случае функционал может зависеть от выбора только одной функции, подчиненной какому-нибудь дополнительному условию).

Связи между функциями могут быть конечные, дифференциальные, интегральные (конечные и дифференциальные связи часто называют обычными связями).

Решение задач на условный экстремум может быть выполнено последовательным исключением функций, но это приводит к значительному усложнению функционала. Такой метод в вариационном исчислении, как правило, не применяется.

В вариационном исчислении решение задач на условный экстремум опирается на теорему, сформулированную ниже для функционалов, зависящих от выбора нескольких функций и их первых производных при конечных связях (эту теорему приведем без доказательства).

Теорема. *Если функции* $y_1, y_2, \ldots, y_n$ *реализуют экстремум функционала*

$$v = \int_{x_1}^{x_2} F(x; y_1, y_2, \ldots, y_n; y_1', y_2', \ldots, y_n')\, dx, \tag{34}$$

$$y_i(x_1) = y_{i1},\ y_i(x_2) = y_{i2} \quad (i = 1, 2, 3, \ldots, n)$$

и связаны условиями

$$\varphi_i(x, y_1, y_2, \ldots, y_n) = 0 \quad (i = 1, 2, \ldots, m, \text{ где } m < n), \tag{35}$$

то они удовлетворяют при соответствующем выборе множителей $\lambda_i(x)$ $(i = 1, 2, \ldots, m)$ *уравнениям Эйлера, составленным для функционала*

$$\bar{v} = \int_{x_1}^{x_2} \left(F + \sum_{i=1}^{m} \lambda_i \varphi_i\right) dx = \int_{x_1}^{x_2} \bar{F}\, dx. \tag{36}$$

При этом предполагается, что уравнения связи независимы.

При решении задачи неизвестные функции $y_i(x)$ и множители $\lambda_i(x)$ находятся из системы уравнений Эйлера

$$\bar{F}_{y_j} - \frac{d}{dx}\bar{F}_{y_j'} = 0 \quad (j = 1, 2, 3, \ldots, n), \tag{37}$$

присоединенных к ней уравнений связи

$$\varphi_i = 0 \quad (i = 1, 2, 3, \ldots, m)$$

и граничных условий.

Множители $\lambda_i(x)$ называются множителями Лагранжа, а метод, основанный на сформулированной теореме,— методом множителей Лагранжа.

Типичной задачей на условный экстремум при конечных связях является задача о геодезических линиях (формулировка этой задачи приведена в § 37).

В случае дифференциальных связей с производными первого порядка (так называемая общая задача Лагранжа) поступают точно также: составляют вспомогательный функционал, применяя, как и в случае конечных связей, метод множителей Лагранжа.

К общей задаче Лагранжа приводятся многие задачи вариационного исчисления. Мы рассмотрим некоторые из них.

Изопериметрическая задача. В настоящее время так называют все задачи с интегральными связями, а не только задачу, упомянутую в § 37.

Решение изопериметрической задачи выполняют, преобразуя интегральные связи в дифференциальные, т. е. приведением к общей задаче Лагранжа. Как это делается, покажем на примере.

Пример. Среди всех гладких кривых L длины l, соединяющих две данные точки A и B ($AB < l$), найти кривую, ограничивающую вместе с отрезком AB наибольшую площадь (рис. 88).

Р е ш е н и е. Прежде всего, выберем систему координат. В качестве оси абсцисс возьмем прямую, проходящую через две данные точки A и B. Пусть при этом абсциссы точек A и B будут соответственно a и b.Ось ординат проведем перпендикулярно к оси абсцисс так, чтобы получилась правая система координат.

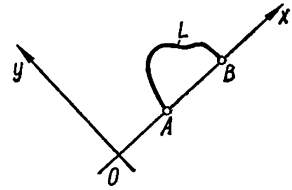

Рис. 88

Запишем уравнение искомой кривой L в виде $y = y(x)$ и предположим для простоты рассуждений, что $y(x)$ — однозначная функция. Тогда площадь, которую ограничит искомая кривая, выразится интегралом:

$$S = \int_a^b y(x)\,dx.$$

Длина дуги этой кривой $y = y(x)$, ограничивающей площадь S, определится формулой

$$\int_a^b \sqrt{1 + y'^2(x)}\,dx = l.$$

Теперь можно сформулировать задачу аналитически так: найти кривую $y = y(x)$, на которой функционал $S(y(x)) = \int_a^b y(x)\,dx$ достигает максимума, если

$$\int_a^b \sqrt{1 + y'^2}\,dx = l,$$

при заданных граничных условиях $y(a) = y(b) = 0$ (не нарушая общности, можно считать, что кривая $y = y(x)$ расположена над осью абсцисс,т. е. $y(x) \geqslant 0$).

Здесь мы имеем изопериметрическую задачу в широком смысле — задачу об отыскании экстремума функционала

$$v(y_i(x)) = \int_{x_1}^{x_2} F(x;\ y_1,\ y_2,\ \ldots,\ y_n;\ y_1',\ y_2',\ \ldots,\ y_n')\,dx \qquad (38)$$

при наличии условий вида

$$\int_{x_1}^{x_2} f_j(x;\ y_1,\ y_2,\ \ldots,\ y_n;\ y_1',\ y_2',\ \ldots,\ y_n')\,dx = l_j \quad (j = 1, 2, \ldots, m) \qquad (39)$$

и при заданных граничных условиях

$$y_i(x_1) = y_{i1},\ y_i(x_2) = y_{i2} \quad (i = 1, 2, 3, \ldots, n). \tag{40}$$

Сведение данной изопериметрической задачи к обычной задаче на условный экстремум выполняем, вводя новую неизвестную функцию. Обозначим

$$\int_a^x \sqrt{1 + y'^2}\,dx = z(x).$$

Тогда $z(a) = 0$ и $z(b) = l$. Дифференцируя $z(x)$ по x, получим

$$z'(x) = \sqrt{1 + y'^2}.$$

Теперь интегральная связь

$$\int_a^b \sqrt{1 + y'^2}\,dx = l$$

предстает в виде дифференциальной связи, т. е. в виде обычных условий для задачи на условный экстремум. Поэтому поставленная задача будет формулироваться уже так: исследовать на максимум функционал

$$S(y(x)) = \int_a^b y(x)\,dx$$

при условии $\sqrt{1 + y'^2} - z'(x) = 0$.

Непосредственному исследованию (согласно приведенной в этом параграфе теореме) подвергаем вспомогательный функционал

$$S^* = \int_a^b (y(x) + \lambda(x)(\sqrt{1 + y'^2} - z'(x)))\,dx = \int_a^b F^*dx,$$

где $F^* = y(x) + \lambda(x)(\sqrt{1 + y'^2} - z'(x))$.

Этот функционал зависит от выбора двух функций $y(x)$ и $z(x)$, поэтому составляем для него два уравнения Эйлера вида (18):

$$F^*_y - \frac{d}{dx}F^*_{y'} = 0 \text{ и } F^*_z - \frac{d}{dx}F^*_{z'} = 0.$$

Так как

$$F^*_y = 1,\ F^*_{y'} = \frac{\lambda(x)\,y'}{\sqrt{1 + y'^2}},\ F^*_z = 0,\ F^*_{z'} = -\lambda(x),$$

то получаем

$$1 + \frac{d}{dx} \cdot \frac{\lambda(x)\,y'}{\sqrt{1 + y'^2}} = 0$$

и

$$\frac{d}{dx}\lambda(x) = 0.$$

Из второго уравнения заключаем, что

$$\lambda(x) = \lambda = \text{const}.$$

Последнее позволяет нам вновь обратиться к вспомогательному функционалу S^* и заметить, что первое из уравнений Эйлера для этого функционала совпадает с уравнением Эйлера для такого более простого функционала:

$$\overline{S} = \int_a^b (y + \lambda \sqrt{1 + y'^2})\, dx = \int_a^b \overline{F} dx,$$

в котором $\overline{F} = y + \lambda \sqrt{1 + y'^2}$ и $\lambda = \text{const}$, а подынтегральная функция $\overline{F}$ не зависит непосредственно от x (см. пятый случай интегрируемости уравнения Эйлера). Поэтому уравнение Эйлера здесь имеет первый интеграл вида

$$\overline{F} - y'\overline{F}_{y'} = C_1.$$

Следовательно,

$$y + \lambda \sqrt{1 + y'^2} - y'\lambda \frac{y'}{\sqrt{1 + y'^2}} = C_1,$$

или

$$y + \frac{\lambda}{\sqrt{1 + y'^2}} = C_1.$$

Теперь введем вспомогательный параметр t по формуле

$$y' = \operatorname{tg} t.$$

Тогда

$$y - C_1 = -\lambda \cos t, \quad dy = \lambda \sin t dt;$$

$$\frac{dy}{dx} = \operatorname{tg} t, \quad dx = \frac{dy}{\operatorname{tg} t} = \frac{\lambda \sin t dt}{\operatorname{tg} t} = \lambda \cos t dt$$

и

$$x = \lambda \sin t + C_2.$$

Итак, экстремали исходного функционала $S(y(x))$ при заданном изопериметрическом (интегральном) условии определяются параметрическими уравнениями:

$$\begin{cases} x - C_2 = \lambda \sin t, \\ y - C_1 = -\lambda \cos t. \end{cases}$$

После исключения параметра t окончательно получим

$$(x - C_2)^2 + (y - C_1)^2 = \lambda^2.$$

Это семейство окружностей, в уравнении которого постоянные C_1, C_2 и λ определяются из условий

$$y(a) = y(b) = 0 \text{ и } \int_a^b \sqrt{1 + y'^2} dx = l.$$

Таким образом, максимум рассмотренного функционала реализуется на дуге окружности длины l, проходящей через точки A и B, наибольшая площадь ограничивается отрезком AB и указанной дугой окружности. Из элементарных геометрических соображений ясно, что в решении получается реализация максимума функционала, но не его минимума.

Следует отметить, что интегральные связи всегда приводят к постоянным значениям множителей Лагранжа. Это позволяет для

изопериметрической задачи сформулировать так называемый принцип взаимности:

экстремали функционала $v = \int_{x_1}^{x_2} F(x, y, y')\,dx$ *при граничных условиях* $y(x_1) = y_1$, $y(x_2) = y_2$ *и интегральной связи* $\int_{x_1}^{x_2} G(x, y, y')\,dx = \text{const}$ *совпадают с экстремалями функционала* $q = \int_{x_1}^{x_2} G(x, y, y')\,dx$ *при граничных условиях* $y(x_1) = y_1$, $y(x_2) = y_2$ *и интегральной связи* $\int_{x_1}^{x_2} F(x, y, y')\,dx = \text{const}$.

Подчеркнем особо, что задачу Лагранжа к изопериметрической в общем случае привести нельзя.

Задача об экстремуме функционала, зависящего от высших производных. Выше было показано, что для такой задачи составляется уравнение Эйлера — Пуассона. Однако с помощью замены переменных названную задачу можно свести к общей задаче Лагранжа.

Например, при нахождении экстремума функционала

$$v = \int_{x_1}^{x_2} F(x, y, y', y'')\,dx \tag{41}$$

можно положить

$$y' = z,\ y'' = z'. \tag{42}$$

Тогда надо решать вариационную задачу для функционала

$$v = \int_{x_1}^{x_2} F(x, y, z, z')\,dx \tag{43}$$

при условии

$$y' - z = 0. \tag{44}$$

А это уже общая задача Лагранжа.

Задача Майера. Одна из форм задачи такова:

Среди систем гладких функций $y_0, y_1, y_2, \ldots, y_n$, удовлетворяющих граничным условиям на одном из концов интервала $[x_1; x_2]$, например, на конце x_1

$$y_0(x_1) = y_{01},\ y_1(x_1) = y_{11},\ \ldots,\ y_n(x_1) = y_{n1} \tag{45}$$

и условиям связи

$$\left.\begin{aligned} &\varphi_i(x, y, y') = 0, \quad i = 0, 1, 2, \ldots, m; \quad m < n) \\ &\varphi_k(x_2, y_0(x_2), y_1(x_2) \ldots, y_n(x_2)) = 0, \quad 0 \leqslant k < n + 1 \end{aligned}\right\} \tag{46}$$

найти ту систему, в которой $y_0(x)$ на втором конце интервала (здесь x_2) имеет экстремум.

Задача Майера приводится к общей задаче Лагранжа.

Примером задачи Майера может служить задача об управлении разгонным двигателем так, чтобы в момент времени T скорость $v(t)$ стала максимально возможной. Приведение этой задачи к общей задаче Лагранжа выполняется с учетом известного факта:

$$v(T) = \int_0^T \dot{v}(t)\,dt, \text{ где } \dot{v}(t) = \frac{dv}{dt}.$$

Задача Больца. Это задача об отыскании экстремума функционала вида

$$v = \int_{x_1}^{x_2} F(x, y, y')\,dx + \Phi(x_1, y(x_1), x_2, y(x_2)) \qquad (47)$$

при наличии уравнений связи и граничных условий.

Задача Больца эквивалентна задаче Лагранжа.

§ 40. ВАРИАЦИОННЫЕ ЗАДАЧИ С ПОДВИЖНЫМИ ГРАНИЦАМИ, С УГЛОВЫМИ ТОЧКАМИ, С ОГРАНИЧЕНИЯМИ

Все рассмотренные выше методы решения вариационных задач относились к случаям, где граничные точки фиксированы и все допустимые кривые проходят через эти точки. Существуют, однако, задачи, в которых одна или обе граничные точки заранее неизвестны и должны выбираться наилучшим образом, исходя из условий минимизации или максимизации рассматриваемого функционала. В некоторых случаях выбор граничных точек стеснен дополнительными требованиями: граничные точки должны принадлежать каким-то кривым или поверхностям, допустимые кривые не должны выходить за пределы какой-то области и т. п. Задачи подобного рода принадлежат к классу вариационных задач с подвижными границами.

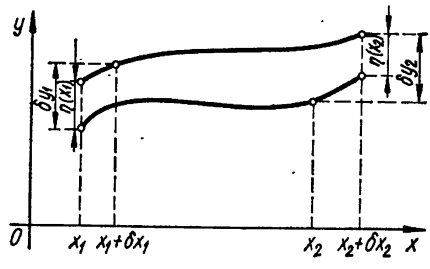

Рис. 89

Общий подход к решению таких задач тот же, что и в случаях фиксированных граничных точек; необходимое условие экстремума функционала — равенство нулю первой вариации — остается в силе. Однако здесь приращение функционала (а, следовательно, и его вариация) возникает не только вследствие перехода на смежную кривую (вариации аргумента), но и в результате смещения (вариации) граничных точек.

Получим первую вариацию простейшего функционала (13)

$$v = \int_{x_1}^{x_2} F(x, y, y')\,dx$$

в случае одновременной вариации аргумента и граничных точек.

Пусть исходная функция — $y(x)$, варьированная — $y(x) + \eta(x)$, абсциссы исходных граничных точек — x_1 и x_2, варьированные абсциссы граничных точек — $x_1 + \delta x_1$ и $x_2 + \delta x_2$ (рис. 89). Тогда приращение функционала

$$\Delta v = \int_{x_1+\delta x_1}^{x_2+\delta x_2} F(x, y+\eta, y'+\eta')\,dx - \int_{x_1}^{x_2} F(x, y, y')\,dx =$$

$$= \int_{x_1}^{x_2} (F(x, y+\eta, y'+\eta') - F(x, y, y'))\,dx +$$

$$+ \int_{x_1}^{x_2+\delta x_2} F(x, y+\eta, y'+\eta')\,dx - \int_{x_1}^{x_1+\delta x_1} F(x, y+\eta, y'+\eta')\,dx.$$

Раскладывая функционал в ряд Маклорена и отбрасывая члены высших порядков, находим

$$\Delta v \approx \int_{x_1}^{x_2} (F_y(x, y, y')\,\eta + F_{y'}(x, y, y')\,\eta')\,dx + F(x, y, y')|_{x=x_2} \cdot \delta x_2 -$$

$$- F(x, y, y')|_{x=x_1} \cdot \delta x_1,$$

откуда, применяя интегрирование по частям и учитывая, что $\eta(x_1) \approx \delta y_1 - y'\delta y_1$, $\eta(x_2) \approx \delta y_2 - y'\delta y_2$, получаем

$$\delta v = \int_{x_1}^{x_2} \left(F_y - \frac{d}{dx} F_{y'}\right)\eta\,dx + F_{y'}|_{x=x_2} \cdot \delta y_2 + (F - F_{y'}y')|_{x=x_2} \cdot \delta x_2 -$$

$$- F_{y'}|_{x=x_1} \cdot \delta y_1 - (F - F_{y'}y')|_{x=x_1} \cdot \delta x_1. \qquad (48)$$

Таким образом, первая вариация функционала состоит из алгебраической суммы интегрального члена, возникающего вследствие вариации аргумента $y(x)$ внутри первоначального интервала интегрирования, и внеинтегральных членов, возникающих вследствие вариации концов интервала.

По предположению граничные точки не закреплены. Их координаты должны выбираться так, чтобы функционал $v(y(x))$ мог достигать экстремума, т. е. чтобы его первая вариация $\delta v(y(x))$ обращалась в нуль.

При этом, если некоторая кривая доставляет экстремум функционалу среди всех допустимых кривых определенного класса, то она тем более будет лучшей среди кривых более узкого класса и проходящих через те же граничные точки. Это значит, что такая кривая $y(x)$ должна быть экстремалью функционала $v(y(x))$,— функция $y(x)$ должна быть решением уравнения Эйлера для этого функционала. Поэтому интегральный член вариации δv обращается в нуль и необходимое условие экстремума функционала приобретает вид

$$F_{y_1}|_{x=x_2} \cdot \delta y_2 + (F - y'F_{y'})|_{x=x_2} \cdot \delta x_2 - F_{y'}|_{x=x_1} \cdot \delta y_1 -$$

$$- (F - y'F_{y'})|_{x=x_1} \cdot \delta x_1 = 0. \qquad (49)$$

Если выбор граничных точек не стеснен никакими дополнительными требованиями, то вариации δx_1, δx_2, δy_1, δy_2 координат этих точек независимы, и потому

$$F_{y'}|_{x=x_2} = 0, \quad (F - y'F_{y'})|_{x=x_2} = 0 \tag{50}$$

$$F_{y'}|_{x=x_1} = 0, \quad (F - y'F_{y'})|_{x=x_1} = 0. \tag{51}$$

Система этих равенств — условия для определения координат граничных точек (x_1, y_1) и (x_2, y_2).

Если есть дополнительные требования, то эти условия соответственно видоизменяются.

Так требование, чтобы граничные точки лежали на некоторых кривых $y = \varphi_1(x)$ и $y = \varphi_2(x)$ (рис. 90)

$$y_2 = \varphi_2(x_2), \tag{52}$$

$$y_1 = \varphi_1(x_1) \tag{53}$$

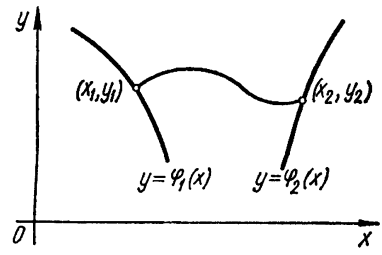

Рис. 90

приводит к зависимости вариаций δx и δy. С точностью до бесконечно малых высших порядков можно записать:

$$\delta y_2 \approx \varphi_2' \delta x_2, \tag{54}$$

$$\delta y_1 \approx \varphi_1' \delta x_1, \tag{55}$$

считая δx_1 и δx_2 произвольными. И тогда из выражения первой вариации получаем

$$(F - F_{y'}(y' - \varphi_2'))|_{x=x_2} = 0 \tag{56}$$

и

$$(F - F_{y'}(y' - \varphi_1'))|_{x=x_1} = 0. \tag{57}$$

Найденные условия (56) и (57) называются условиями трансверсальности. Они устанавливают связь между угловыми коэффициентами $\varphi'(x)$ и $y'(x)$ в граничных точках. Совместно с условиями (52) и (53) они определяют постоянные интегрирования в решениях уравнений Эйлера и, вообще говоря, дают возможность найти не только положения концов экстремали, но и ее наклон в этих концах по отношению к заданным кривым.

Если свободна только одна граничная точка, например (x_2, y_2), а другая фиксирована, то координаты фиксированной точки (здесь x_1 и y_1) не варьируются, их вариации равны нулю ($\delta x_1 = \delta y_1 = 0$), и для определения постоянных интегрирования используются граничные условия в фиксированной точке (x_1, y_1) и условия трансверсальности в свободной точке.

Для функционалов вида

$$v = \int_{x_1}^{x_2} f(x, y)\sqrt{1 + y^2}\,dx \tag{58}$$

условия трансверсальности сводятся к ортогональности экстремали и заданной кривой. Действительно,

$$F_{y'} = f(x, y)\frac{y'}{\sqrt{1+y'^2}},$$

поэтому условие трансверсальности, например, в точке (x_2, y_2) будет

$$\frac{F}{1+y'^2}(1+y'\varphi_2') = 0,$$

откуда при $F \neq 0$

$$y' = -\frac{1}{\varphi_2'(x)}. \tag{59}$$

Для функционалов вида

$$v = \int_{x_1}^{x_2} F(x, y, z, y', z')\,dx, \tag{60}$$

зависящих от выбора двух функций, условия трансверсальности находятся подобно простейшему случаю. При независимых вариациях δx, δy, δz они записываются так:

$$(F - y'F_{y'} - zF_{z'})|_{x=x_i} = 0,\ F_{y'}|_{x=x_i} = F_{z'}|_{x=x_i} = 0$$
$$(i = 1, 2).$$

Если граничная точка должна находиться на пространственной кривой

$$\begin{cases} y = \varphi(x), \\ z = \psi(x), \end{cases} \tag{62}$$

причем

$$y_1 = \varphi(x_1) = \varphi_1,\ y_2 = \varphi(x_2) = \varphi_2,$$
$$z_1 = \psi(x_1) = \psi_1,\ z_2 = \psi(x_2) = \psi_2, \tag{63}$$

то

$$(F + (\varphi' - y')F_{y'} + (\psi' - z')F_{z'})|_{x=x_i} = 0 \quad (i = 1, 2). \tag{64}$$

Для функционала (41), зависящего от первой и второй производных

$$v = \int_{x_1}^{x_2} F(x, y, y', y'')\,dx,$$

имеют место следующие условия трансверсальности при независимых вариациях δx_1, δx_2, δy_1, δy_2, $\delta y_1'$, $\delta y_2'$:

$$\begin{cases} (F - y'F_{y'} - y''F_{y''})|_{x=x_i} = 0, \\ \left(F_{y'} - \frac{d}{dx}F_{y''}\right)\Big|_{x=x_i} = 0, \\ F_{y''}|_{x=x_i} = 0 \quad (i = 1, 2). \end{cases} \tag{65}$$

Для более сложных функционалов и при сложных связях между вариациями координат граничных точек условия трансверсальности существенно усложняются.

В ряде случаев экстремум функционала достигается не на непрерывных и гладких кривых. Экстремали бывают кусочно-гладкими, с угловыми точками. Рассмотрим такой случай для простейшего функционала.

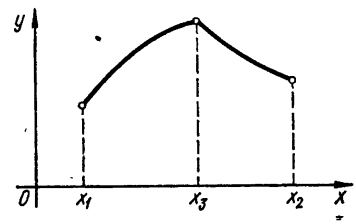

Рис. 91

Для простоты исследования предположим, что кривая, доставляющая экстремум функционалу (13)

$$v = \int_{x_1}^{x_2} F(x, y, y')\,dx,$$

при граничных условиях $y(x_1) = y_1$, $y(x_2) = y_2$ имеет единственный излом в точке x_3, находящейся между x_1 и x_2 (рис. 91).

На основании свойств интегралов можно записать

$$v = \int_{x_1}^{x_2} F(x, y, y')\,dx + \int_{x_3}^{x_2} F(x, y, y')\,dx, \tag{66}$$

т. е. представить заданный функционал в виде суммы двух функционалов. Назовем

$$v_1 = \int_{x_1}^{x_3} F(x, y, y')\,dx, \quad v_2 = \int_{x_3}^{} F(x, y, y')\,dx, \tag{67}$$

тогда $v = v_1 + v_2$, и в соответствии с определением первой вариации

$$\delta v = \delta v_1 + \delta v_2. \tag{68}$$

При нахождении экстремалей положение точки излома заранее неизвестно, поэтому, если обозначить $y(x_3) = y_3$, то

$$\left.\begin{aligned} \delta v_1 &= F_{y_1}|_{x=x_3-0} \cdot \delta y_3 + (F - y'F_{y'})|_{x=x_3-0} \cdot \delta x_3, \\ \delta v_2 &= -F_{y'}|_{x=x_3+0} \cdot \delta y_3 - (F - y'F_{y'})|_{x=x_3+0} \cdot \delta x_3, \end{aligned}\right\} \tag{69}$$

поскольку точки (x_1, y_1) и (x_2, y_2) фиксированы.

Необходимым условием экстремума функционала является равенство нулю его первой вариации:

$$\delta v = \delta v_1 + \delta v_2 = 0,$$

или

$$\begin{aligned} &F_{y'}|_{x=x_3-0} \cdot \delta y_3 + (F - y'F_{y'})|_{x=x_3-0} \cdot \delta x_3 - \\ &- F_{y'}|_{x=x_3+0} \cdot \delta y_3 - (F - y'F_{y'})|_{x=x_3+0} \cdot \delta x_3 = 0, \end{aligned} \tag{70}$$

т. е.

$$\begin{aligned} &(F_{y'}|_{x=x_3-0} - F_{y'}|_{x=x_3+0}) \cdot \delta y_3 + ((F - y'F_{y'})|_{x=x_3-0} - \\ &- (F - y'F_{y'})|_{x=x_3+0}) \cdot \delta x_3 = 0. \end{aligned} \tag{71}$$

Отсюда, вследствие произвольности δx_3 и δy_3 получаем, что

$$\left.\begin{aligned} F_{y'}|_{x=x_3-0} &= F_{y'}|_{x=x_3+0}, \\ (F - y'F_{y'})|_{x=x_3-0} &= (F - y'F_{y'})|_{x=x_3+0}. \end{aligned}\right\} \qquad (72)$$

Одновременно на каждом из участков экстремали выполняется уравнение Эйлера (18).

Условия (72) называются условиями Вейерштрасса — Эрдмана. Они позволяют определить недостающие произвольные постоянные в решениях уравнения Эйлера — в уравнениях экстремалей.

Кривая, составленная из решений уравнения Эйлера таким образом, что в точках излома выполняются условия Вейерштрасса — Эрдмана, называется ломаной экстремалью.

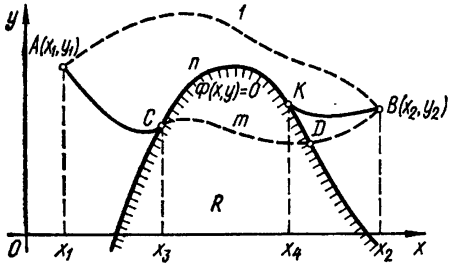

Рис. 92

Для задач об оптимальном управлении характерно наличие ограничений. Ознакомимся с тем, что в таких задачах получается, на примере простейшего функционала (13)

$$v = \int_{x_1}^{x_2} F(x, y, y')\,dx$$

при граничных условиях $y(x_1) = y_1$, $y(x_2) = y_2$.

Пусть на рис. 92 кривая $\Phi(x, y) = 0$ (или $y = \varphi(x)$ — после разрешения уравнения относительно y) ограничивает область R, в которую не должны заходить допустимые кривые, соединяющие две фиксированные граничные точки $A(x_1, y_1)$ и $B(x_2, y_2)$.

Среди решений $y = y(x)$ уравнения Эйлера (18) для заданного функционала (13) могут оказаться целиком проходящие вне области R (на рисунке — кривая *1*), это будут обычные экстремали, а могут оказаться и кривые, пересекающие область R (на рисунке — кривая $ACmDB$).

Тогда экстремум функционала не может реализоваться на всей экстремали. В таких случаях экстремум функционала может достигаться на составной линии (ее мы назовем псевдоэкстремалью), состоящей из части AC экстремали, проходящей через точку A до граничной линии $\Phi(x, y) = 0$, участка CnK граничной линии и части KB некоторой экстремали, проходящей через точку B. Этот участок может и не принадлежать первоначальной экстремали $ACmDB$.

Таким образом, экстремаль начала процесса своей точкой C скользит по границе $\Phi(x, y) = 0$, экстремаль конца процесса скользит по той же линии своим концом K. Этим рассматриваемая задача напоминает случай подвижных концов экстремали; здесь в точках C и K имеют смысл условия трансверсальности. Однако, кроме того, часть CnK граничной кривой сама превращается в дугу, на

которой может реализоваться экстремум функционала. Такое положение приводит к тому, что условия трансверсальности в точках C и K превращаются в условия гладкого сопряжения.

В итоге интервал интегрирования $[x_1; x_2]$ в функционале v разбивается на три части. Поэтому, если обозначить черех x_3 и x_4 координаты подвижных концов (соответственно точек C и K), то функционал v предстанет как сумма трех функционалов:

$$v = \int_{x_1}^{x_3} F(x, y, y')\,dx + \int_{x_3}^{x_4} F(x, y, y')\,dx + \int_{x_4}^{x_2} F(x, y, y')\,dx, \tag{73}$$

или

$$v = v_1 + v_2 + v_3, \tag{74}$$

где

$$v_1 = \int_{x_1}^{x_3} F(x, y, y')\,dx, \quad v_2 = \int_{x_3}^{x_4} F(x, y, y')\,dx,$$

$$v_3 = \int_{x_4}^{x_2} F(x, y, y')\,dx. \tag{75}$$

Приравнивая к нулю суммарную вариацию

$$\delta v = \delta v_1 + \delta v_2 + \delta v_3, \tag{76}$$

после необходимых преобразований получаем

$$y'(x_3) = \varphi'(x_3) \tag{77}$$

и

$$y'(x_4) = \varphi'(x_4), \tag{78}$$

что возможно при условиях

$$F_{y'y'}|_{x=x_3} = 0, \tag{79}$$

$$F_{y'y'}|_{x=x_4} = 0. \tag{80}$$

Таким образом, имеет место гладкое сопряжение. При этом искомая псевдоэкстремаль, состоящая из частей двух экстремалей и заключенного между ними участка граничной кривой, полностью определяется шестью параметрами — двумя постоянными интегрирования для экстремали начала процесса, двумя постоянными интегрирования для экстремали конца процесса и двумя абсциссами x_3 и x_4 точек сопряжения. Для нахождения этих параметров есть шесть условий — граничные условия в начальной и в конечной точках (по одному условию в каждой точке: $y(x_1) = y_1$ и $y(x_2) = y_2$) и четыре условия сопряжения — по два в каждой точке:

$$y(x_3) = \varphi(x_3), \quad y'(x_3) = \varphi'(x_3), \tag{81}$$

$$y(x_4) = \varphi(x_4), \quad y'(x_4) = \varphi'(x_4). \tag{82}$$

Если одна из граничных точек интервала лежит на ограничивающей кривой, то псевдоэкстремаль состоит только из одного участка экстремали и участка ограничивающей кривой, а условия сопряжения выполняются только в одной точке.

Пример. Найти линию, на которой может реализоваться экстремум функционала

$$v(y(x)) = \int_{-2}^{3} (y + y'^2)\, dx,$$

если $y(-2) = y(3) = 1$ при условии, что область допустимых кривых ограничена требованием $y \geqslant 2x - x^2$.

Р е ш е н и е. Уравнение Эйлера для заданного функционала имеет вид

$$1 - 2y'' = 0.$$

Его общее решение (уравнение семейства экстремалей)

$$y = \frac{x^2}{4} + C_1 x + C_2.$$

Необходимо найти экстремали, проходящие через заданные граничные точки, и найти (если они существуют) две точки x_3 и x_4 сопряжения найденных экстремалей с ограничивающей кривой $y = 2x - x^2$, поскольку ни одна из граничных точек не лежит на ограничивающей кривой.

Для дальнейшего обозначим

$$\varphi(x) = 2x - x^2.$$

В левой части графика имеем

$$y(-2) = 1 = 1 - 2C_1 + C_2, \quad \text{или} \quad C_2 = 2C_1.$$

Условия сопряжения экстремали с ограничивающей кривой (в точке x_3)

$$\begin{cases} \varphi(x_3) = y(x_3), \\ \varphi'(x_3) = y'(x_3). \end{cases}$$

Поэтому для отыскания точки x_3 получаем систему уравнений

$$\begin{cases} \dfrac{x^2}{4} + C_1 x + C_2 = 2x - x^2, \\ \dfrac{x}{2} + C_1 = 2 - 2x, \\ 2C_1 = C_2. \end{cases}$$

Из этой системы имеем

$$C_2 = 2C_1, \quad C_1 = 2 - \frac{5}{2}x, \quad x = \frac{-10 \pm 6\sqrt{5}}{5}.$$

Так как $\dfrac{-10 - 6\sqrt{5}}{5}$ не входит в интервал интегрирования, то

$$x_3 = \frac{-10 + 6\sqrt{5}}{5} = -2 + \frac{6}{\sqrt{5}},$$

а потому

$$C_1 = 7 - 3\sqrt{5}, \quad C_2 = 14 - 6\sqrt{5}.$$

Уравнение левой экстремали имеет вид

$$y_1 = \frac{x^2}{4} + (7 - 3\sqrt{5})\, x + (14 - 6\sqrt{5}).$$

При этом в точке сопряжения

$$y_1(x_3) = \varphi(x_3) = \frac{-76}{5} + \frac{36}{\sqrt{5}},$$

$$y_1'(x_3) = \varphi'(x_3) = 6 - \frac{12}{\sqrt{5}}.$$

В правой части графика

$$y(3) = 1 = \frac{9}{4} + 3C_1 + C_2, \quad C_2 = -\frac{5}{4} - 3C_1.$$

Условия сопряжения граничной кривой и экстремали в точке x_4 записываются в виде

$$\begin{cases} \dfrac{x^2}{4} + C_1 x + C_2 = 2x - x^2, \\ \dfrac{x}{2} + C_1 = 2 - 2x, \\ C_2 = -\dfrac{5}{4} - 3C_1. \end{cases}$$

Из этой системы имеем

$$C_2 = -\frac{5}{4} - 3C_1, \quad C_1 = 2 - \frac{5}{2}x, \quad x = \frac{15 \pm 4\sqrt{5}}{5}.$$

Поскольку $x = \dfrac{15 + 4\sqrt{5}}{5}$ не входит в интервал интегрирования, то

$$x_4 = \frac{15 - 4\sqrt{5}}{5} = 3 - \frac{4}{\sqrt{5}}.$$

Из этого следует

$$C_1 = \frac{-11 + 4\sqrt{5}}{2}, \quad C_2 = \frac{61 - 24\sqrt{5}}{4}.$$

Уравнение правой экстремали имеет вид

$$y_2 = \frac{x^2}{4} - \frac{11 - 4\sqrt{5}}{2}x + \frac{61 - 24\sqrt{5}}{4}.$$

При этом в точке сопряжения

$$y_2(x_4) = \varphi(x_4) = \frac{-13 + 16\sqrt{5}}{5},$$

$$y_2'(x_4) = \varphi'(x_4) = -4 + \frac{8}{\sqrt{5}}.$$

Таким образом, экстремум функционала может быть достигнут на псевдоэкстремали, состоящей из трех частей:

$$y = \begin{cases} \dfrac{x^2}{4} + (7 - 3\sqrt{5})x + (14 - 6\sqrt{5}) & \text{при } -2 \leqslant x \leqslant x_3, \\ 2x - x^2 & \text{при } x_3 \leqslant x \leqslant x_4, \\ \dfrac{x^2}{4} - \dfrac{11 - 4\sqrt{5}}{2}x + \dfrac{61 - 24\sqrt{5}}{4} & \text{при } x_4 \leqslant x \leqslant 3, \end{cases}$$

при этом в точках сопряжения

$$x_3 = -2 + \frac{6}{\sqrt{5}}, \quad y(x_3) = \frac{-76}{5} + \frac{36}{\sqrt{5}}, \quad y'(x_3) = 6 - \frac{12}{\sqrt{5}},$$

$$x_4 = 3 - \frac{4}{\sqrt{5}}, \quad y(x_4) = \frac{-31 + 16\sqrt{5}}{5}, \quad y'(x_4) = -4 + \frac{8}{\sqrt{5}}.$$

§ 41. ПОНЯТИЕ О ДОСТАТОЧНЫХ УСЛОВИЯХ ЭКСТРЕМУМА ФУНКЦИОНАЛА

Все рассмотренное выше ограничивалось лишь необходимыми условиями экстремума функционала.

Выяснение достаточных условий экстремума связано с более сложными исследованиями, в некоторой степени напоминающими исследование на экстремум функции многих переменных в математическом анализе. Там исследование функции нескольких переменных требует прежде всего введения в рассмотрение второго дифференциала, причем необходимым условием экстремума является равенство нулю первого дифференциала, а достаточным — знакоопределенность второго дифференциала при одновременном обращении в нуль первого дифференциала.

Для исследования достаточных условий экстремума функционала используется вторая вариация. Покажем, как это делается для простейшего функционала (13)

$$v = \int_{x_1}^{x_2} F(x, y, y')\, dx.$$

В § 38 было дано определение второй вариации этого функционала с помощью ряда Маклорена. Вернемся к этому определению.

Предположим, что $F(x, y, y')$ — трижды дифференцируемая функция своих аргументов. Пусть $y = y(x)$ — непрерывная дифференцируемая функция, на которой реализуется экстремум функционала (13) и пусть далее $y = y(x) + \alpha\eta(x)$ — некоторая смежная функция, причем $\eta(x)$ — непрерывно дифференцируемая функция и $\eta(x_0) = \eta(x_1) = 0$; α — малый параметр.

Значение функционала на смежной кривой, очевидно, равно

$$\int_{x_1}^{x_2} F(x, y + \alpha\eta, y' + \alpha\eta')\, dx, \tag{83}$$

а приращение функционала будет

$$\Delta v(y(x)) = \int_{x_1}^{x_2} (F(x, y + \alpha\eta, y' + \alpha\eta') - F(x, y, y'))\, dx. \tag{84}$$

По формуле Маклорена имеем

$$F(x, y + \alpha\eta, y' + \alpha\eta') = F(x, y, y') + \alpha(F_y(x, y, y') +$$

$$+ F_{y'}(x, y, y')) + \frac{1}{2}\alpha^2 (F_{yy}(x, y, y')\eta^2 + 2F_{yy'}(x, y, y')\eta\eta' +$$

$$+ F_{y'y'}(x, y, y')\eta'^2) + \alpha^3 r_2(\alpha), \tag{85}$$

где $\alpha^3 r_2(\alpha)$ — бесконечно малые высших порядков.
Так как

$$\Delta v = \delta v + \frac{1}{2!}\delta^2 v + \alpha^3 r_2(\alpha), \tag{11}$$

то теперь легко получаем явные выражения для первой и второй вариаций функционала (13):

$$\delta v = \alpha \int_{x_1}^{x_2} (\eta F_y(x, y, y') + \eta' F_{y'}(x, y, y'))\, dx, \tag{86}$$

$$\delta^2 v = \alpha^2 \int_{x_1}^{x_2} (\eta^2 F_{yy}(x, y, y') + 2\eta\eta' F_{yy'}(x, y, y') +$$

$$+ \eta'^2 F_{y'y'}(x, y, y'))\, dx. \tag{87}$$

Из разложения (11) приходим к уже известному заключению, что необходимым условием существования экстремума функционала является равенство нулю его первой вариации. Из этого же разложения следует, что достаточным условием экстремума функционала должно быть знакопостоянство на интервале $[x_1; x_2]$ второй вариации, когда его первая вариация равна нулю, т. е. знакопостоянство второй вариации функционала вдоль рассматриваемой экстремали на интервале $[x_1; x_2]$.

Таким образом, при $\delta v = 0$ должен сохраняться знак интеграла (87)

$$\int_{x_1}^{x_2} (\eta^2 F_{yy}(x, y, y') + 2\eta\eta' F_{yy'}(x, y, y') + \eta'^2 F_{y'y'}(x, y, y'))\, dx.$$

Однако непосредственно установить условия знакопостоянства такого интеграла весьма затруднительно. Кроме того, достаточные условия слабого и сильного экстремумов различны. Поэтому приходится выполнить ряд преобразований подынтегральной функции, ввести в рассмотрение некоторые новые необходимые условия экстремума функционала и лишь после этого получить достаточные условия экстремума.

Для изучения достаточных условий экстремума простейшего функционала (13)

$$v(y(x)) = \int_{x_1}^{x_2} F(x, y, y')\, dx$$

остановимся на следующих понятиях и фактах.

Понятие поля. В теории обыкновенных дифференциальных уравнений для интегральных кривых дифференциальных уравнений

первого порядка было введено понятие п о л я н а п р а в л е н и й. Уточним это понятие, учитывая, что в вариационном исчислении рассматриваются дифференциальные уравнения второго порядка.

Семейство кривых $y = y(x, C)$ в области D образует поле (точнее собственное поле), если через каждую точку области D проходит одна и только одна кривая этого семейства.

Угловой коэффициент $k(x, y)$ касательной к кривой, образующей поле, называется н а к л о н о м поля в точке (x, y).

Например, внутри прямоугольника $a \leqslant x \leqslant b$, $c \leqslant y \leqslant d$, который не содержит начала координат, семейство концентрических окружностей $x^2 + y^2 = C^2$ образует поле с переменным наклоном $k(x, y) = -\frac{x}{y}$, а семейство парабол $y = (x - C)^2 - 2$ внутри того же прямоугольника поля не образует, так как параболы указанного семейства пересекаются внутри прямоугольника, т. е. через точки выбранной области уже проходит не по одной, а более кривых семейства.

Если все кривые семейства $y = y(x, C)$ проходят через одну и ту же точку (x_0, y_0) области D, то поля в смысле указанного выше определения они не образуют. Однако, если во всех точках области, кроме центра пучка (x_0, y_0), требования, предъявляемые к полю, выполнены, то и здесь считается, что семейство кривых $y = y(x, C)$ образует поле, которое, в отличие от собственного поля, называют ц е н т р а л ь н ы м.

Если поле образовано экстремалями какой-либо вариационной задачи, то оно называется п о л е м э к с т р е м а л е й.

Если какая-нибудь экстремаль входит в семейство, образующее поле экстремалей, то говорят, что эта экстремаль включена в поле (или что поле включает экстремаль).

Различные формы второй вариации функционала. Уравнение Якоби. Необходимые условия Лежандра и Якоби для экстремума функционала. Допустим, что $y = \varphi(x)$ является экстремалью функционала, т. е. эта функция удовлетворяет уравнению Эйлера и граничным условиям $\varphi(x_1) = y_1$, $\varphi(x_2) = y_2$.

Тогда, так как $\delta v = 0$, приращение функционала имеет вид

$$\Delta v = \frac{1}{2}\delta^2 v + \alpha^3 r_2(\alpha), \tag{88}$$

причем (в силу малости слагаемого $\alpha^3 r_2(\alpha)$) знак этого приращения полностью определяется знаком второй вариации функционала

$$\delta^2 v = \alpha^2 \int_{x_1}^{x_2} (\eta^2 F_{yy}(x, \varphi(x), \varphi'(x)) + 2\eta\eta' F_{yy'}(x, \varphi(x), \varphi'(x)) + \eta'^2 F_{y'y'}(x, \varphi(x), \varphi'(x)))\, dx \tag{89}$$

(здесь учтено, что вторая вариация рассматривается на экстремали $y = \varphi(x)$).

Обозначим

$$\left.\begin{aligned} F_{yy}(x, \varphi, \varphi') &= P(x), \\ F_{yy'}(x, \varphi, \varphi') &= Q(x), \\ F_{y'y'}(x, \varphi, \varphi') &= R(x). \end{aligned}\right\} \tag{90}$$

Тогда вторая вариация функционала запишется так:

$$\delta^2 v = \alpha^2 \int_{x_1}^{x_2} (P(x)\eta^2 + 2Q(x)\eta\eta' + R(x)\eta'^2)\, dx. \tag{91}$$

Первая форма второй вариации (Якоби). Если обозначить $2\Omega = P\eta^2 + 2Q\eta\eta' + R\eta'^2$, то можно записать

$$2\Omega = \eta(P\eta + Q\eta') + \eta'(Q\eta + R\eta').$$

Замечая, что

$$P\eta + Q\eta' = \frac{\partial \Omega}{\partial \eta} \quad \text{и} \quad Q\eta + R\eta' = \frac{\partial \Omega}{\partial \eta'},$$

уравнение (91) запишем в виде

$$\delta^2 v = \alpha^2 \int_{x_1}^{x_2} \left(\eta \frac{\partial \Omega}{\partial \eta} + \eta' \frac{\partial \Omega}{\partial \eta'}\right) dx.$$

Интегрируя второе слагаемое этого уравнения по частям, получим

$$\delta^2 v = \alpha^2 \int_{x_1}^{x_2} \eta(x) \left(\frac{\partial \Omega}{\partial \eta} - \frac{d}{dx} \frac{\partial \Omega}{\partial \eta'}\right) dx.$$

Обозначим полные производные:

$$\frac{\partial Q}{dx} = Q', \quad \frac{\partial \eta}{\partial x} = \eta', \quad \frac{\partial (R\eta')}{\partial x} = (R\eta')'$$

тогда

$$\delta^2 v = \alpha^2 \int_{x_1}^{x_2} \eta(x) ((P - Q')\eta - (R\eta')')\, dx. \tag{92}$$

Зависимость (92) — это и есть первая форма второй вариации функционала.

Учитывая состав полученного выражения (92), рассмотрим дифференциальное уравнение

$$(P - Q')z - (Rz')' = 0 \tag{93}$$

или в развернутом виде

$$Rz'' + R'z' + (Q' - P)z = 0. \tag{94}$$

Отметим, что дифференциальное уравнение (93) получается по основной лемме вариационного исчисления, если в (92) положить $\delta^2 v = 0$ и вместо η и η' записать соответственно z и z'.

Возвращаясь к первоначальным обозначениям, можно вместо (93) записать

$$\left(F_{yy} - \frac{d}{dx} F_{yy'}\right) z - \frac{d}{dx} (F_{y'y'} z') = 0$$

и вместо (94) соответственно

$$F_{yy} z'' + \frac{dF_{y'y'}}{dx} z' + \left(\frac{d}{dx} F_{yy'} - F_{yy}\right) z = 0.$$

Уравнение (94) называется дифференциальным уравнением Якоби.

Легко заметить, что если уравнение Якоби имеет непрерывное, равное нулю на всем отрезке $[x_1; x_2]$ решение, то $\delta^2 v = 0$, и по второй вариации в общем случае выводов об экстремуме функционала делать нельзя. Если же $\delta^2 v = 0$ и $\delta^3 v \neq 0$, то этого достаточно для отсутствия экстремума функционала.

Вторая форма второй вариации. Предполагая что $z = z(x)$ есть решение уравнения Якоби, отличное от нуля на всем интервале $[x_1; x_2]$, и учитывая что

$$(\eta R z')' = \eta (R z')' + R z' \eta', \quad (z R \eta')' = z (R \eta')' + R \eta' z',$$
$$\eta (x_1) = \eta (x_2) = 0,$$

можно сравнительно легко получить форму второй вариации в виде

$$\delta^2 v = \alpha^2 \int_{x_1}^{x_2} \frac{R(x)}{z^2} \cdot \begin{vmatrix} \eta' & \eta \\ z' & z \end{vmatrix}^2 dx. \tag{95}$$

Из этого выражения следует, что знак второй вариации зависит от знака $R(x)$.

Если $R(x)$ будет сохранять свой знак, то можно подозревать наличие экстремума. Разумеется, что в этом случае функция $R(x)$ не должна иметь корней внутри интервала $[x_1; x_2]$.

Третья форма второй вариации. Интегрируя по частям второе слагаемое первоначальной формы второй вариации (91)

$$\delta^2 v = \alpha^2 \int_{x_1}^{x_2} (P \eta^2 + 2Q \eta \eta' + R \eta'^2)\, dx,$$

получаем

$$\int_{x_1}^{x_2} 2Q \eta \eta' dx = \int_{x_1}^{x_2} Q d(\eta^2) = - \int_{x_1}^{x_2} \eta^2 Q' dx,$$

а потому

$$\delta^2 v = \alpha^2 \int_{x_1}^{x_2} ((P - Q') \eta^2 + R \eta'^2)\, dx. \tag{96}$$

Эта зависимость является третьей формой второй вариации.

Из уравнения (96) следует (с учетом выводов из второй формы вариации), что для существования экстремума функционала необ-

ходимо, чтобы на экстремали функция $R(x) = F_{yy'}$ сохраняла свой знак неизменным ($F_{yy'} \geqslant 0$ — необходимое условие Лежандра для минимума функционала, уже известное нам из § 38).

Можно показать, что условие $F_{y'y'} > 0$ (усиленное условие Лежандра) не всегда будет достаточным для наличия минимума функционала.

Рассмотрим еще одно необходимое условие экстремума функционала. Изучим условия, при которых решение уравнения Якоби (94) не обращается в нуль ни при каких значениях x на отрезке $[x_1; x_2]$.

Если $R(x) \neq 0$ и сохраняет знак на интервале $[x_1; x_2]$, то уравнение Якоби можно записать так:

$$z'' + \frac{R'}{R} z' + \frac{Q' - P}{R} z = 0. \tag{97}$$

Коэффициенты этого уравнения $\frac{R'}{R}$ и $\frac{Q' - P}{R}$ непрерывны и дифференцируемы на интервале $[x_1; x_2]$. Поэтому все решения уравнения Якоби также непрерывны и дифференцируемы на этом интервале. Обозначим через $z_1(x)$ и $z_2(x)$ два линейно-независимых решения уравнения (97). Тогда

$$z_1'' + \frac{R'}{R} z_1' + \frac{Q' - P}{R} z_1 \equiv 0,$$

$$z_2'' + \frac{R'}{R} z_2' + \frac{Q' - P}{R} z_2 \equiv 0.$$

Умножая первое из этих тождеств на z_2, второе — на z_1 и почленно вычитая, получаем

$$z_1 z_2'' - z_1'' z_2 + \frac{R'}{R}(z_1 z_2' - z_1' z_2) = 0,$$

или

$$\frac{z_1 z_2'' - z_1'' z_2}{z_1 z_2' - z_1' z_2} + \frac{R'}{R} = 0,$$

так как вследствие линейной независимости решений

$$z_1 z_2' - z_1' z_2 \neq 0.$$

После умножения обеих частей полученного равенства на dx и интегрирования найдем, что

$$(z_1 z_2' - z_1' z_2) R = C. \tag{98}$$

Отсюда следует:

а) два непрерывных линейно-независимых решения уравнения Якоби не могут иметь одинаковых корней на интервале $[x_1; x_2]$;

б) каждый корень решения простой ($z_1(x)$ и $z_1'(x)$ одновременно в нуль не обращаются, так же как $z_2(x)$ и $z_2'(x)$);

в) между двумя последующими корнями одного решения лежит корень (и притом только один) второго решения, линейно-независимого от первого (теорема Штурма).

Теперь общее решение уравнения Якоби можно записать в виде

$$z(x) = C_1 z_1(x) + C_2 z_2(x), \tag{99}$$

где C_1 и C_2 произвольные постоянные.

Подберем C_1 и C_2 таким образом, чтобы $z(x_1) = 0$.

Точка $\tilde{x}$ называется сопряженной с точкой x_1, если выбранное решение $z(x)$ уравнения Якоби, не равное тождественно нулю, обращается в нуль также при $x = \tilde{x}$, т. е. если

$$z(\tilde{x}) = z(x_1) = 0.$$

Нетрудно убедиться в том, что если вторая вариация $\delta^2 v$ не изменяет своего знака (при выполнении условия Лежандра) и не обращается в нуль, то на интервале $[x_1; x_2]$ нет точек, сопряженных с x_1. Благодаря этому приходим к следующему утверждению:

Для того чтобы функционал $\int_{x_1}^{x_2} F(x, y, y')\,dx$ *достигал на экстремали* $y = y(x)$ *экстремума, необходимо, чтобы вдоль этой экстремали на интервале* $[x_1; x_2]$ *не было точек, сопряженных с точкой* x_1 *(необходимое условие Якоби).*

К сказанному добавим еще два определения точки, сопряженной с x_1.

а) Точка $\tilde{x}$ называется сопряженной с точкой $x = x_1$, если в ней разность между данной экстремалью $y(x)$ и произвольно близкой экстремалью $\tilde{y}(x)$, выходящей из той же начальной точки, есть величина выше первого порядка малости по сравнению с $|\tilde{y}(x) - y(x)|$ (т. е. по сравнению с модулем разности между этими экстремалями в любой другой точке).

б) Точка $\tilde{x}$ называется сопряженной с точкой $x = x_1$, если она есть предел точек пересечения данной экстремали $y(x)$ близкими экстремалями, выходящими из той же начальной точки, при $\tilde{y}(x) - y(x) \to 0$.

Отметим, что эти два определения сопряженной точки указывают на то, что для отсутствия вдоль экстремали на интервале $[x_1; x_2]$ корней решений уравнения Якоби необходимо включение рассматриваемой экстремали в поле экстремалей.

Справедливо и другое утверждение:

Для построения поля экстремалей с центром в точке (x_1, y_1) *и содержащего дугу данной экстремали достаточно, чтобы на этой дуге не было точек, сопряженных с точкой* x_1.

Функция Вейерштрасса. Пусть дан простейший функционал (13)

$$v = \int_{x_1}^{x_2} F(x, y, y')\,dx$$

при граничных условиях $y(x_1) = y_1$, $y(x_2) = y_2$ и пусть рассматриваемая экстремаль $y = y(x)$ (обозначим ее через L) включена в поле экстремалей с наклоном поля $p(x, y)$. На этой экстремали подынтегральная функция будет $F(x, y, p)$. Меняя значение y', будем изменять значение функции $F(x, y, y')$, одновременно переходя на смежную экстремаль. Очевидно, в таком случае частное приращение по y' функции $F(x, y, y')$

$$\Delta_{y'} F(x, y, y')$$

допускает представление с помощью частного дифференциала и бесконечно малых высших порядков:

$$\Delta_{y'} F(x, y, p) = F(x, y, y') - F(x, y, p) = \\ = d_{y'} F(x, y, p) + \beta\Delta(y'), \tag{100}$$

где $\Delta(y')$ — приращение производной y' и $\beta \to 0$ при $\Delta(y') \to 0$. Так как

$$d_{y'} F(x, y, p) = (y' - p) F_p(x, y, p), \tag{101}$$

то, обозначая

$$\beta\Delta(y') = E(x, y, p, y'), \tag{102}$$

получим

$$E(x, y, p, y') = F(x, y, y') - F(x, y, p) - \\ - (y' - p) F_p(x, y, p). \tag{103}$$

Функция $E(x, y, p, y')$ называется функцией Вейерштрасса.

Можно показать, что при переходе от экстремали L к некоторой другой близкой допустимой кривой $\tilde{L}$ из того же поля приращение функционала выражается формулой

$$\Delta v = \int_{x_1}^{x_2} E(x, y, p, y')\, dx. \tag{104}$$

Из этого следует, что знак приращения функционала определяется знаком функции Вейерштрасса; отсюда ясно, что неотрицательности функции Вейерштрасса на экстремали L, включенной в поле, достаточно для наличия на этой экстремали минимума функционала, а неположительности функции Вейерштрасса — его максимума.

Этот вывод называется условием Вейерштрасса. Можно показать, что условие Вейерштрасса является одновременно и необходимым для наличия экстремума функционала.

Подводя итог всем рассмотренным в этом параграфе вопросам, заключаем, что для достижения простейшим функционалом $v(y(x))$ экстремума на кривой L достаточно одновременного выполнения следующих условий.

Для слабого экстремума:

а) Кривая L входит в семейство экстремалей, т. е. является интегральной кривой уравнения Эйлера и при этом удовлетворяет граничным условиям.

б) Экстремаль L может быть включена в поле экстремалей (или удовлетворяет условию Якоби).

в) Функция Вейерштрасса не меняет знака во всех точках (x, y), близких к кривой L при близких к p значениях y'.

При этом, если $E \geqslant 0$, имеет место минимум, если $E \leqslant 0$,— максимум.

Для сильного экстремума:

Условия а) и б) остаются теми же, условие в) ослабляется и гласит: функция Вейерштрасса не меняет знака во всех точках (x, y), близких к кривой L, и при произвольных значениях y'. Можно показать, что

$$E(x, y, p, y') = \frac{(y' - p)^2}{2!} F_{y'y'}(x, y, q), \qquad (105)$$

где q — некоторое значение, заключенное между p и y'.

Из этого следует, что функция Вейерштрасса сохраняет свой знак, если не изменяет знака $F_{y'y'}(x, y, q)$ в точках, близких к экстремали, и при q, близких к p. Такое положение позволяет использовать для исследования достаточных условий экстремума также и условия Лежандра.

Окончательно получаем три варианта достаточных условий слабого экстремума и три варианта достаточных условий сильного экстремума простейшего функционала.

Так, слабый минимум функционала $v = \int_{x_1}^{x_2} F(x, y, y')\, dx$ достигается на кривой L, уравнение которой $y = y(x)$, если для нее одновременно выполняются условия одного из следующих вариантов:

Вариант 1

а) $F_y - \frac{d}{dx} F_{y'} = 0$; б) условие Якоби; в) $F_{y'y'} > 0$ на исследуемой экстремали.

Вариант 2

а) $F_y - \frac{d}{dx} F_{y'} = 0$; б) условие Якоби; в) $E(x, y, p, y') \geqslant 0$ для точек (x, y), близких к точкам исследуемой экстремали, и для y', близких к $p(x, y)$.

Вариант 3

а) $F_y - \frac{d}{dx} F_{y'} = 0$; б) существует поле экстремалей, включающее исследуемую экстремаль; в) $E(x, y, p, y') \geqslant 0$ для точек (x, y), близких к точкам исследуемой экстремали, и для y', близких к $p(x, y)$.

Для сильного минимума в каждом из трех вариантов условия а) и б) не изменяются, а условие в) ослабляется, а именно: в первом варианте допускается $F_{y'y'} \geqslant 0$, а во втором и в третьем вариантах

исключается требование близости y' к значению $p\ (x, y)$. (Для точек (x, y), близких к точкам на исследуемой экстремали и для произвольных значений y'. Предполагается, что $F\ (x, y, y')$ трижды дифференцируема по y' для малых y').

Достаточные условия максимума (слабого и сильного) отличаются от сформулированных только знаками неравенств.

§ 42. НЕКОТОРЫЕ ПРЯМЫЕ МЕТОДЫ ВАРИАЦИОННОГО ИСЧИСЛЕНИЯ

Дифференциальные уравнения вариационного исчисления решаются в конечном виде чрезвычайно редко. Поэтому появляется необходимость использовать приближенные методы, которые в вариационном исчислении называются прямыми методами. Основная идея прямых методов состоит в следующем.

Функционал $v\ (y\ (x))$ можно рассматривать, как функцию бесконечного числа переменных. Действительно, если допустимые функции раскладываются в ряды (по степенным, тригонометрическим или произвольным функциям:

$$y\ (x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx,$$

$$y\ (x) = \sum_{n=0}^{\infty} a_n \varphi_n\ (x)),$$

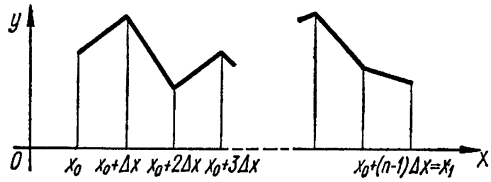

Рис. 93

то для определения функции $y\ (x)$ (а значит и функционала $v\ (y\ (x))$) достаточно задать бесконечную последовательность коэффициентов $a_0, a_1, a_2, \ldots, a_n, \ldots$. Поэтому функционал $v\ (y\ (x))$ оказывается функцией бесконечного множества переменных $a_0, a_1, a_2, \ldots, a_n, \ldots$:

$$v\ (y\ (x)) = \varphi\ (a_0, a_1, a_2, \ldots, a_n, \ldots).$$

Таким образом, вариационное исчисление исследует на экстремум функции бесконечного множества переменных, в отличие от задач на экстремум для функций конечного числа переменных, которые решаются методами дифференциального исчисления.

В прямых методах вариационного исчисления вариационная задача рассматривается как предельная задача на экстремум функции конечного числа переменных.

Существует несколько прямых методов. Рассмотрим здесь только три из них.

Конечно-разностный метод Эйлера. В этом методе значения функционала

$$v\ (y\ (x)) = \int_{x_0}^{x_1} F\ (x, y, y')\, dx \qquad (106)$$

рассматриваются не на всех (произвольных) допустимых кривых, а только на ломаных, составленных из произвольно выбранного

числа n прямолинейных звеньев, с известными абсциссами вершин:

$$x_0,\ x_0+\Delta x,\ x_0+2\Delta x,\ \ldots,\ x_0+k\Delta x,\ \ldots,\ x_0+(n-1)\,\Delta x,$$

где $\Delta x=\frac{x_1-x_0}{n}$ (рис. 93).

В таком случае исследуемый функционал становится функцией $n-1$ переменных — ординат вершин ломаной (концы ломаной закреплены: $y(x_0)=Y_0$, $y(x_1)=Y_1$):

$$v(y(x))=\varphi(y_1,\ y_2,\ \ldots,\ y_{n-1}), \tag{107}$$

где

$$y_1=y(x_0+\Delta x),\quad y_2=y(x_0+2\Delta x),\quad \ldots,\quad y_{n-1}= \\ =y(x_0+(n-1)\,\Delta x). \tag{108}$$

Выберем ординаты $y_1, y_2, \ldots, y_{n-1}$ так, чтобы функция $\varphi(y_1, y_2, \ldots, y_{n-1})$ имела экстремум, а затем перейдем к пределу при $n\to\infty$. Ограничиваясь необходимым условием экстремума, надо, как известно, подобрать $y_1, y_2, \ldots, y_{n-1}$ так, чтобы выполнялось равенство нулю частных производных:

$$\frac{\partial\varphi}{\partial y_1}=\frac{\partial\varphi}{\partial y_2}=\ldots=\frac{\partial\varphi}{\partial y_{n-1}}=0.$$

В пределе (когда $n\to\infty$) при некоторых ограничениях, налагаемых на функцию $F(x, y, y')$, мы получим решение вариационной задачи.

Однако значения функционала на указанных ломаных проще вычислять приближенно (по формуле прямоугольников), заменяя в подынтегральной функции значения производной y' с помощью конечных разностей:

$$\int_{x_0}^{x_1}F(x,\ y,\ y')\,dx\approx\sum_{k=0}^{n-1}\int_{x_0+k\Delta x}^{x_0+(k+1)\Delta x}F\left(x,\ y,\ \frac{y_{k+1}-y_k}{\Delta x}\right)dx\approx \\ \approx\sum_{k=1}^{n}F\left(x_k,\ y_k,\ \frac{\Delta y_k}{\Delta x}\right)\Delta x. \tag{109}$$

Метод Ритца. Идея метода Ритца состоит в том, что функционал $v(y(x))$ рассматривается не на всех произвольных допустимых кривых, а лишь на всевозможных линейных комбинациях первых n функций специально выбранной последовательности $W_i(x)$ ($i=1, 2, \ldots, n$):

$$y_n=\sum_{i=1}^{n}a_iW_i(x), \tag{110}$$

где все $a_i=\text{const}$, а $W_i(x)$ называются координатными функциями.

Так как функции y_n должны быть допустимыми в конкретной задаче, то в этой задаче получается ограничение на выбор последо-

вательности функций $W_i(x)$. Учитывая что $W_i(x)$ — это функции выбранной последовательности, т. е. функции известные, замечаем, что функционал оказывается функцией n коэффициентов:

$$v(y(x)) = \varphi(a_1, a_2, \ldots, a_n). \tag{111}$$

Коэффициенты a_k выбираются так, чтобы функция $\varphi(a_1, a_2, \ldots, a_n)$ имела экстремум. Необходимым условием для этого будет

$$\frac{\partial \varphi}{\partial a_k} = 0 \quad (k = 1, 2, 3, \ldots, n). \tag{112}$$

Выражение $y_n = \sum_{i=1}^{n} a_i W_i$ при выбранных по условиям $\frac{\partial \varphi}{\partial a_k} = 0$ коэффициентах a_k ($k = 1, 2, \ldots, n$) и соблюдении граничных условий определит приближенно экстремаль функционала, т. е. даст решение вариационной задачи. Часто можно ограничиться этим приближенным решением.

Если же приближенного решения недостаточно, то, совершая предельный переход при $n \to \infty$, получают уточненное решение вариационной задачи.

При выборе координатных функций исходят из условий конкретной задачи. Здесь надо требовать, чтобы выполнялись граничные условия; но нельзя забывать и о других условиях (непрерывности, гладкости и т. д.).

Пример. Применяя метод Ритца, найти экстремаль функционала

$$v(y(x)) = \int_0^1 (y'^2 - y^2 - 2xy)\, dx$$

при условиях $y(0) = y(1) = 0$.

Р е ш е н и е. Строим систему координатных функций по формулам: $W_0(x) = 0$, $W_1(x) = x - x^2$, $W_2(x) = x^2 - x^3$, ..., $W_n(x) = x^n - x^{n+1}$, ..., однако в этом конкретном случае считаем приближенно, ограничивая n.

При $n = 2$ имеем

$$y_2(x) = a_1(x - x^2) + a_2(x^2 - x^3),$$

$$y_2'(x) = a_1(1 - 2x) + a_2(2x - 3x^2),$$

$$v(y_2(x)) = \varphi(a_1, a_2) = \int_0^1 ((a_1(1 - 2x) + a_2(2x - 3x^2))^2 - (a_1(x - x^2) +$$

$$+ a_2(x^2 - x^3))^2 - 2x(a_1(x - x^2) + a_2(x^2 - x^3)))\, dx = \frac{3}{10} a_1^2 + \frac{3}{10} a_1 a_2 +$$

$$+ \frac{13}{105} a_2^2 - \frac{1}{6} a_1 - \frac{1}{10} a_2.$$

Итак,

$$\varphi(a_1, a_2) = \frac{3}{10} a_1^2 + \frac{3}{10} a_1 a_2 + \frac{13}{105} a_2^2 - \frac{1}{6} a_1 - \frac{1}{10} a_2.$$

Используя условия $\frac{\partial \Phi}{\partial a_1} = 0$ и $\frac{\partial \Phi}{\partial a_2} = 0$, получим

$$\begin{cases} \frac{3}{5} a_1 + \frac{3}{10} a_2 = \frac{1}{6}, \\ \frac{3}{10} a_1 + \frac{26}{105} a_2 = \frac{1}{10}, \end{cases}$$

откуда $a_1 = \frac{71}{369}$, $a_2 = \frac{7}{41}$, поэтому

$$y_2(x) = \frac{-9x^3 - 8x^2 + 71x}{369}.$$

Метод Канторовича. Этот метод применяется к функционалам, зависящим от выбора одной функции двух переменных

$$v(z(x, y)) = \iint_D F\left(x, y, z(x, y), \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}, \frac{\partial^2 z}{\partial x^2}, \frac{\partial^2 z}{\partial x \partial y}, \frac{\partial^2 z}{\partial y^2}, \ldots\right) dx dy. \quad (113)$$

Решение вариационной задачи находят в виде конечного ряда по выбранной системе координатных функций, которые зависят от двух переменных, а коэффициентами ряда считаются неизвестные функции одной переменной, например, x:

$$z_n(x, y) = \sum_{k=1}^{n} W_k(x, y) f_k(x) + W_0(x, y). \quad (114)$$

Так как $W_k(x, y)$ ($k = 0, 1, 2, 3, \ldots, n$) известны, то в итоге, после подстановки $z_n(x, y)$ в (113) и интегрирования по y, получается функционал, зависящий от выбора n функций $f_k(x)$ ($k = 1, 2, 3, \ldots, n$), уже только одной переменной. Этот новый функционал исследуется на экстремум, что и дает приближенное решение первоначальной вариационной задачи.

При необходимости получения точного решения во всех трех рассмотренных методах выполняется предельный переход при $n \to \infty$.

§ 43. ПОНЯТИЕ О ПРИНЦИПЕ МАКСИМУМА

Рассмотренные выше классические методы вариационного исчисления для многих практических задач нахождения оптимальных процессов и решений при ограничениях переменных в виде неравенств оказываются малоэффективными или вообще непригодными. Последнее относится, например, к случаям, когда требуется найти положение наибольшего (рис. 94, точка *1*) или наименьшего (рис. 94, точка *4*) значения функции или функционала на границе области изменения переменных, а также при наличии двустороннего экстремума неаналитического типа (рис. 94, точка *3*). Подобную геометрическую интерпретацию имеет большинство задач оптимизации, связанных с наложением ограничений на управление или на коорди-

наты. Так, в частности, может быть сформулирована задача движения летательного аппарата, при котором предполагается ограниченное отклонение рулей управления (с ограничением на управляющие координаты). Это происходит потому, что аэродинамические поверхности аппарата обладают различной несущей (подъемной) силой на разных высотах (сказывается плотность атмосферы). Если оптимальный полет выполняется на определенной высоте и управление им физически обеспечивается, то при необходимости изменения высоты полета в сторону ее увеличения (по метеорологическим причинам) рулей может «не хватить» для управления, и по этой причине полет может оказаться не оптимальным (затянут во времени, с перерасходом горючего и т. п.). При этом следует иметь в виду, что усилители мощности, используемые в автоматических системах летательных аппаратов, имеют линейные характеристики лишь до ограниченной сверху величины входного сигнала (при больших сигналах наступает так называемое насыщение). Здесь, как видно, становится проблемой иная, чем прежде, постановка вариационной задачи — как выбрать из числа возможных траекторий полета ту, которая обеспечит наибольший выигрыш при ограничении управления? Для решения подобных вариационных задач оптимизации с ограничениями на координаты объекта или на управляющие координаты регулятора потребовалась разработка новых, отличных от классических, методов решения.

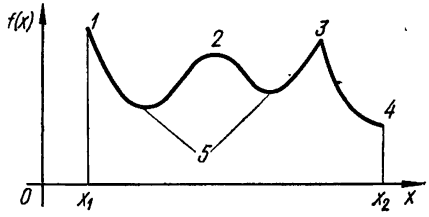

Рис. 94

Одним из таких методов решения задач оптимального управления является изящный метод акад. Л. С. Понтрягина и его учеников[1], получивший название принципа максимума. Перед рассмотрением сущности этого метода остановимся кратко на постановке общих задач оптимального управления.

Пусть движение некоторой динамической системы характеризуется дифференциальным уравнением

$$\frac{dx}{dt} = f(x, u), \tag{115}$$

где x — вектор-столбец состояния движения (см. гл. VI),
u — вектор-столбец управления:

$$x = \begin{Vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Vmatrix} \quad u = \begin{Vmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{Vmatrix}.$$

[1] См.: Л. С. Понтрягин, В. Г. Болтянский, Р. В. Гамкрелидзе, Е. Ф. Мищенко. Математическая теория оптимальных процессов. М., Физматгиз, 1961.

Вектор управления подчинен дополнительно условию

$$g(u) \leqslant 0. \tag{116}$$

При заданном ограничении могут возникнуть три постановки задачи, отличающиеся целевой функцией (см. гл. VIII).

Первая задача состоит в реализации движения за кратчайшее время $\min\limits_{u} \int\limits_{t_0}^{t} dt$ (задача Майера).

Вторая задача заключается в минимизации (максимизации) некоторого функционала в конечный момент времени. Иными словами, к конечному моменту времени t_k при поставленной задаче $F(x(t_k)) = A$ отклонение от A должно быть минимальным.

Третья задача требует, чтобы с учетом ограничения (116) управление u доставляло максимум (минимум) интегралу

$$J(u) = \int\limits_{t_0}^{t_k} F(x, u, t)\, dt. \tag{117}$$

Подобным интегралом могут оцениваться ресурс, надежность, стоимость, расход топлива и др.

Академик Л. С. Понтрягин предложил основной критерий

$$\rho = \sum_{i=1}^{n} b_i x_i(t_k) \quad (\rho > 0). \tag{118}$$

Суть этого критерия можно пояснить так.

Пусть b означает вектор-столбец желаемого движения, а $x(t_k)$ — вектор движения в конечный момент времени. Тогда минимум ρ (близость к нулю) означает приближенную ортогональность векторов x и b — истинное движение максимально приближается к желаемому. Конечная точка оптимальной траектории t_k может быть свободной или закрепленной. Аналогично вектор b может быть жестко ориентирован в пространстве или изменяемым.

Одновременное удовлетворение начальным и граничным условиям осложняет задачу вариационного исчисления. Например, если при постановке задачи оптимального управления требуется максимизировать (или минимизировать) значение интеграла от квадрата одной из переменных, то вводя новую переменную

$$x_{n+1} = \int\limits_{t_0}^{t_k} x_s^2\, dt,$$

получают дополнительное нелинейное уравнение

$$\frac{dx_{n+1}}{dt} = x_s^2, \quad x_{n+1}(0) = 0.$$

Заслуга акад. Л. С. Понтрягина состоит в том, что он предположил существование неразрывной связи оптимального управления с максимумом энергии и сформулировал это как специальный прин-

цип — принцип максимума. Этот принцип гласит, что если вектор $\boldsymbol{u}$ оптимален, т. е. если он обеспечивает минимум функции Понтрягина ρ, то энергетическая функция Гамильтона $H(\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{u}, t)$ имеет максимум по отношению к $\boldsymbol{u}$ в интервале управления.

Функция Гамильтона

$$H(\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{u}, t) = (\boldsymbol{p} \cdot \boldsymbol{f}) = \sum_{j=1}^{n} p_j f_j, \tag{119}$$

где $\boldsymbol{p}$ — вектор количества движения, $\boldsymbol{f}$ — вектор правой части уравнения (115).

Вектор $\boldsymbol{p}$ определяется как решение дифференциального уравнения

$$\frac{dp_i}{dt} = -\sum_{j=1}^{n} p_j \frac{\partial f_j}{\partial x_i} \quad (i = 1, 2, \ldots, n), \tag{120}$$

где $p_i(t_k) = -b_i$, b_i — некоторые известные постоянные, входящие в функцию Понтрягина.

Дифференцируя уравнение (119) по p_i, найдем

$$\frac{\partial H}{\partial p_i} = f_i. \tag{121}$$

С другой стороны, дифференцирование этого же уравнения (119) по x_i дает

$$\frac{\partial H}{\partial x_i} = \sum_{j=1}^{n} p_j \frac{\partial f_j}{\partial x_i} \quad (i = 1, 2, \ldots, n). \tag{122}$$

Учитывая (115) и (120), получаем

$$\left.\begin{aligned} \frac{\partial x_i}{dt} &= \frac{dH}{\partial p_i}, \\ \frac{dp_i}{dt} &= -\frac{dH}{dx_i} \end{aligned}\right\} \tag{123}$$

— канонические уравнения Гамильтона.

Эти уравнения подчинены начальным и граничным условиям:

$$\left.\begin{aligned} x_i(t_0) &= x_i^0 \quad (i = 1, 2, \ldots, n) \\ p_i(t_k) &= -b_i. \end{aligned}\right\} \tag{124}$$

Поскольку $\boldsymbol{p}$ — вектор количества движения, а $\boldsymbol{f}$, согласно уравнению (115), также пропорциональна $\frac{d\boldsymbol{x}}{dt}$, то их скалярное произведение, равное H (см. (119)), пропорционально энергии.

Остается показать, что максимум H совпадает с минимумом ρ, и тогда открывается прямой путь к решению задачи оптимального управления через решение канонической системы уравнений (123) при условиях (124). Придавая приращение функции Понтрягина и принимая во внимание, что

$$p_i(t_k) = -b_i,$$

получим

$$\Delta\rho = \sum_{i=1}^{n} b_i \Delta x_i \Big|_{t_0}^{t_k} = -\sum_{i=1}^{n} p_i \Delta x_i \Big|_{t_0}^{t_k}.$$

Здесь учтено, что $\Delta x_i\,(t_0) = 0$.

Продолжая подстановку Δx_i из (115), находим, что

$$\Delta\rho = -\int_{t_0}^{t_k} \sum_{i=0}^{n} p_i \left(f_i\,(x^* + \Delta x,\ u^* + \Delta u,\ t) - f_i\,(x^*,\ u^*,\ t)\right) dt =$$

$$= -\int_{t_0}^{t_k} \sum_{i=1}^{n} p_i \left(f_i\,(x^*,\ u^* + \Delta u,\ t) - f_i\,(x^*,\ u^*,\ t) + \right.$$

$$+ \sum_{j=1}^{n} \frac{\partial f_i\,(x^*,\ u^* + \Delta u,\ t)}{\partial x_j} \Delta x_j + \sum_{s=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f_i\,(x^* + \xi\Delta x,\ u^* + \Delta u,\ t)}{\partial x_j \partial x_s} \times$$

$$\left. \times \Delta x_j \Delta x_s + \ldots \right) dt.$$

В этом разложении через x^* и u^* обозначены оптимальные значения x и u; коэффициент $0 < \xi < 1$, и для обоюдного приращения x и u далее применена формула Тейлора.

Рассмотрим частный случай задания правой части f_i, а именно:

$$f_i = \sum_{j=1}^{n} a_{ij}\,(t)\, x_j + \varphi\,(u,\ t).$$

Здесь управление и координаты разделены, что характерно для линейных автоматических систем.

В таком случае слагаемые с $\frac{\partial f_i}{\partial x_j}$ и $\frac{\partial^2 f_i}{\partial x_j \partial x_s}$ исчезнут и приращение

$$\Delta\rho = -\int_{t_0}^{t_k} \left(H\,(x^*,\ u^* + \Delta u,\ t) - H\,(x^*,\ u^*,\ t)\right) dt. \qquad (125)$$

Отсюда видно, что минимум ρ ($\Delta\rho \geqslant 0$) получается при

$$H\,(x^*,\ u^* + \Delta u,\ t) - H\,(x^*,\ u^*,\ t) \leqslant 0.$$

В случае нелинейной функции f_i это же положение Понтрягина доказывается сложнее [1].

Процедура решения задачи оптимального уравнения по Понтрягину сводится к следующему:

1. По исходной системе уравнений (115) составляется сопряженная система (120) и далее функция H (119).

[1] См.: сборник «Методы оптимизации с приложениями к механике космического полета». М., Наука, Главная редакция физико-математической литературы, 1965, с. 311—313.

2. Отыскивается максимум H по u. Осуществляется запись канонических уравнений, начальных и граничных условий.

3. Подставляя условия максимума функции Гамильтона в канонические уравнения, находят дифференциальные уравнения, определяющие оптимальный процесс при граничных условиях.

Примеры

1. Дано уравнение $\frac{d^2x}{dt^2} = u$ и $|u| < u_m$.

Решение. Обозначим $x = x_1$ и $\frac{dx}{dt} = x_2$. Тогда получим

$$\frac{dx_1}{dt} = x_2, \quad \frac{dx_2}{dt} = u.$$

Функция Гамильтона $H = (p \cdot f) = p_1x_2 + p_2u$. Сопряженная система $\frac{dp_1}{dt} = 0$, $\frac{dp_2}{dt} = -p_1$. Отсюда $p_1 = C_1$, $p_2 = C_2 - C_1t$ и функция $H = C_1x_2 + (C_2 - C_1t)\,u$. $H_{\max}$ будет при положительном слагаемом $(C_2 - C_1t)\,u$, т. е. при $u = u_m \operatorname{sign}(C_2 - C_1t)$.

Функция $(C_2 - C_1t)$ дает уравнение прямой линии, пересекающей ось t один раз. Следовательно, оптимальная по быстродействию система оказывается релейной, а управление имеет два предельных значения $+u_m$ и $-u_m$.

2. Имеется объект с уравнением

$$T\frac{dx}{dt} + x = ku, \quad |u| \leqslant u_m,$$

при начальном значении $x(t_0) = x_0$.

Определить управляющее воздействие u так, чтобы обеспечить минимум квадратичной оценки

$$J(u) = \int_{t_0}^{t_k} x^2(u, t)\,dt.$$

Решение. Пусть $x = x_1$, а $x_2 = \int_{t_0}^{t_k} x_1^2 dt$. Тогда каноническая форма уравнений приобретает вид

$$\begin{cases} \frac{dx_1}{dt} = -\frac{1}{T}\,x + \frac{k}{T}\,u, \\ \frac{dx_2}{dt} = x_1^2. \end{cases}$$

Функция Понтрягина $\rho = b_1x_1 + b_2x_2$. Определим u^*, доставляющее минимум ρ.

Функция Гамильтона

$$H = p_1f_1 + p_2f_2 = p_1\left(-\frac{1}{T}\,x_1 + \frac{k}{T}\,u\right) + p_2x_1^2,$$

или

$$H = p_2\left(x_1 - \frac{p_1}{2p_2T}\right)^2 - \frac{p_1^2}{4p_2T^2} + \frac{p_1k}{T}u.$$

Отсюда видно, что u будет доставлять H максимум, если совпадают знаки p_1 и u, т. е. $u = u_m \operatorname{sign} p_1$.

Канонические уравнения Гамильтона имеют вид

$$\frac{dx_i}{dt} = \frac{\partial H}{\partial p_i}\begin{cases} \dfrac{dx_1}{dt} = -\dfrac{1}{T}x + \dfrac{k}{T}u_m \operatorname{sign} p_1, \\ \dfrac{dx_2}{dt} = x_1^2; \end{cases}$$

$$\frac{\partial p_i}{dt} = -\frac{\partial H}{\partial x_i}\begin{cases} \dfrac{dp_1}{dt} = \dfrac{p_1}{T} - 2x_1p_2, \\ \dfrac{dp_2}{dt} = 0. \end{cases}$$

Начальные условия для x будут $x_1(t_0) = x_0$, $x_2(t_0) = 0$. Граничные условия $p_1(t_k) = -b_1$, $p_2(t_k) = -b_2$. Учитывая, что $\dfrac{dp_2}{dt} = 0$, получаем $p_2 = \text{const}$.

$$\frac{dp_1}{dt} = \frac{p_1}{T} + 2x_1b_2.$$

Теперь необходимо решить уравнения

$$\begin{cases} T\dfrac{dx_1}{dt} + x_1 = ku_m \operatorname{sign} p_1, \\ T\dfrac{dp_1}{dt} - p_1 = 2Tb_2x_1 \end{cases}$$

при $x_1(t) = x_0$ и $p_1(t_k) = -b_1$.

Если $b_1 > 0$, то $\operatorname{sign} p_1(t_k)$ отрицательный, и в конечной точке после решения уравнения $T\dfrac{dx_1}{dt} + x_1 = -ku_m$ получаем

$$x_1 = -ku_m\left(1 - e^{-\frac{t_k}{T}}\right).$$

С другой стороны, задавая $x_1 = x_0$, после решения уравнения

$$T\frac{dp_1}{dt} - p_1 = 2Tb_2x_0$$

находим

$$p_1 = 2Tb_2x_0\left(1 - e^{\frac{t}{T}}\right).$$

Если $t = 0$, то $p_1 = 0$, а если $t > 0$, то $p_1 < 0$. Как оптимальный процесс можно принять решение в виде

$$x = -ku_m\left(1 - e^{-\frac{t}{T}}\right).$$

ЛИНЕЙНОЕ И ДИНАМИЧЕСКОЕ ПРОГРАММИРОВАНИЕ

Линейное и динамическое программирования представляют собой особые математические дисциплины, позволяющие решать задачи рационального (оптимального) планирования процессов.

В задачах линейного программирования решение сводится к нахождению наибольшего (или наименьшего) значения некоторой линейной функции при определенных ограничениях, заданных в виде системы линейных неравенств.

В задачах динамического программирования осуществляется поэтапное планирование многошагового процесса, при котором на каждом этапе оптимизируется только один шаг. При этом управление на каждом шаге выбирается, исходя из эффективности планирования в целом.

Линейное и динамическое программирования широко применяются при решении задач экономики (управления, рационального планирования и организации производства). Они находят все большее применение также в инженерной практике при выборе рациональных параметров технических устройств и, в частности, при создании различных систем автоматического регулирования.

§ 44. ЗАДАЧА ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Для того чтобы дать математическую формулировку задачи линейного программирования в общем виде, предварительно рассмотрим несколько примеров.

1. Транспортная задача. Пусть имеются три угольных бассейна с фиксированными объемами добычи в (млн. т) и четыре пункта потребления с известными объемами потребления угля (в млн. т)

Таблица 5

Угольний бассейн	1	2	3	Всего
Объем добычи	48	48	40	136

Таблица 6

Пункт потребления	1	2	3	4	Всего
Объем потребления	32	57	32	15	136

В табл. 7 указаны затраты, связанные с транспортировкой груза от угольных бассейнов к пунктам потребления (в сотнях тыс. рублей).

Требуется найти наиболее целесообразное закрепление пунктов потребления за угольными бассейнами с точки зрения экономии затрат по транспортировке.

Таблица 7

Угольный бассейн	Пункт потребления			
	1	2	3	4
1	5	8	24	30
2	8	10	20	32
3	17	28	35	40

Обозначим через x_{ij} количество угля, которое с i-го ($i = 1, 2, 3$) угольного бассейна получает j-й ($j = 1, 2, 3, 4$) пункт потребления.

Тогда система уравнений

$$\begin{cases} x_{11} + x_{12} + x_{13} + x_{14} = 48, \\ x_{21} + x_{22} + x_{23} + x_{24} = 48, \\ x_{31} + x_{32} + x_{33} + x_{34} = 40 \end{cases} \tag{1}$$

характеризует количество угля, вывозимого с каждого угольного бассейна, а система уравнений

$$\begin{cases} x_{11} + x_{21} + x_{31} = 32, \\ x_{12} + x_{22} + x_{32} = 57, \\ x_{13} + x_{23} + x_{33} = 32, \\ x_{14} + x_{24} + x_{34} = 15 \end{cases} \tag{2}$$

определяет количество угля, получаемое каждым пунктом потребления.

Затраты на перевозку груза:

в первый пункт потребления составляют $f_1 = 5x_{11} + 8x_{21} + 17x_{31}$,

во второй пункт потребления составляют $f_2 = 8x_{12} + 10x_{22} + 28x_{32}$,

в третий пункт потребления составляют $f_3 = 24x_{13} + 20x_{23} + 35x_{33}$,

в четвертый пункт потребления составляют $f_4 = 30x_{14} + 32x_{24} + 40x_{34}$.

Суммарные затраты F будут

$$F = f_1 + f_2 + f_3 + f_4 = 5x_{11} + 8x_{21} + 17x_{31} + 8x_{12} + 10x_{22} + 28x_{32} + 24x_{13} + 20x_{23} + 35x_{33} + 30x_{14} + 32x_{24} + 40x_{34}. \tag{3}$$

После выяснения этих вопросов можно дать следующую математическую постановку задачи. Найти неотрицательные решения системы линейных уравнений (1) при выполнении условий (2) так, чтобы линейная форма F (3) принимала наименьшее значение (минимизировалась). (Ясно, что можно находить и решение системы уравнений (2) при выполнении условий (1)).

Таблица 8

Виды сырья	Запасы сырья	Виды продукции	
		$П_1$	$П_2$
S_1	19	2	3
S_2	13	2	1
S_3	15	0	3
S_4	18	3	0
Доход		7	5

2. Задача об использовании сырья. Пусть изготовление двух видов продукции $П_1$ и $П_2$ требует использования четырех видов сырья S_1, S_2, S_3, S_4. Запасы каждого вида сырья на производстве ограничены. При этом запасы и количества каждого вида сырья, необходимого для производства (в условных единицах) каждого вида продукции, указаны в табл. 8. В последней строке приведен доход (в рублях) предприятия от реализации одной единицы каждого вида продукции.

Требуется составить такой план выпуска продукции, при котором доход предприятия оказался бы максимальным.

Здесь математическая постановка задачи получается следующим путем.

Если принять, что предприятие выпускает x_1 единиц продукции вида $П_1$ и x_2 — вида $П_2$, то для этого требуется сырья:

$$\begin{aligned} S_1 &— 2x_1 + 3x_2 \text{ единиц,} \\ S_2 &— 2x_1 + x_2 \text{ единиц,} \\ S_3 &— 3x_2 \text{ единиц,} \\ S_4 &— 3x_1 \text{ единиц.} \end{aligned}$$

Очевидно, расход сырья не должен превышать имеющихся запасов, т. е.

$$\begin{cases} 2x_1 + 3x_2 \leqslant 19, \\ 2x_1 + x_2 \leqslant 13, \\ 3x_2 \leqslant 15, \\ 3x_1 \leqslant 18. \end{cases} \tag{4}$$

При этих условиях доход предприятия составит

$$F = 7x_1 + 5x_2 \text{ рублей.} \tag{5}$$

Теперь задачу можно сформулировать так:

Дана система неравенств (4) и линейная форма (5). Требуется среди неотрицательных решений системы линейных неравенств (4) найти такое, при котором линейная форма F (5) имеет наибольшее значение (максимизируется).

3. Задача об использовании оборудования. Пусть производству в соответствии с планом по номенклатуре требуется изготовить 50 единиц продукции вида $П_1$, 30 единиц продукции вида $П_2$ и 45 единиц продукции вида $П_3$. Каждый из видов продукции может изготовляться на двух машинах A_1 и A_2, причем производительность этих

машин (в единицах времени, необходимого для производства одной единицы продукции данного вида на данной машине) указана в табл. 9. Требуется составить план использования оборудования, т. е. указать количество x_{ij} продукции Π_j ($j = 1, 2, 3$), которое следует изготовлять на машине A_i ($i = 1, 2$), с тем чтобы задание по номенклатуре оказалось выполненным.

Эта задача заключается в отыскании оптимального плана использования оборудования, т. е. плана, при котором задание будет выполнено в минимальное время. Она может получить математическую формулировку следующим образом.

Таблица 9

Машина	Продукция		
	Π_1	Π_2	Π_3
A_1	4	10	10
A_2	6	8	20

Таблица 10

A_i	Π_j		
	Π_1	Π_2	Π_3
A_1	x_{11}	x_{12}	x_{13}
A_2	x_{21}	x_{22}	x_{23}
План изготовления	50	30	45

Искомые величины x_{ij} для удобства сведем в табл. 10.

Очевидно, согласно требованию выполнения плана по номенклатуре, должно быть

$$\begin{cases} x_{11} + x_{21} = 50, \\ x_{12} + x_{22} = 30, \\ x_{13} + x_{23} = 45. \end{cases} \tag{6}$$

Так как известно время, необходимое каждой машине для изготовления одной единицы каждого вида продукции (табл. 9), то, учитывая принятые обозначения и табл. 10, можно определить время t_1 и соответственно t_2, необходимое первой и второй машинам для изготовления намеченного количества продукции и выполнения плана по номенклатуре.

Имеем

$$\begin{cases} t_1 = 4x_{11} + 10x_{12} + 10x_{13}, \\ t_2 = 6x_{21} + 8x_{22} + 20x_{23}. \end{cases} \tag{7}$$

Машины работают одновременно, поэтому общее время T, необходимое для выполнения плана, равно наибольшему из значений t_1 и t_2, т. е.

$$T = \max(t_1, t_2). \tag{8}$$

Теперь математическая формулировка задачи будет такой:

Найти такие неотрицательные решения системы уравнений (6), при которых величина (8) достигает своего наименьшего значения.

4. Задача проектирования системы автоматического управления[1]. Одной из актуальных задач любого инженерного проектирования

[1] См.: А. М. Бонч-Бруевич, В. Л. Быков, П. И. Чинаев. Бесконтактные элементы самонастраивающихся схем. М., Машиностроение, 1968, с. 40—42.

является задача обеспечения необходимого запаса надежности конструкции без превышения при этом заданных технологических и стоимостных характеристик.

Пусть требуется спроектировать систему автоматического управления так, чтобы ее надежность R была наибольшей, а масса, габариты и стоимость меньше некоторых наперед заданных величин G^*, V^* и C^* соответственно. Параметры каждого узла системы: надежность r_i, масса g_i, габариты v_i и стоимость c_i известны.

Надежность системы с последовательным (основным) соединением n различных узлов

$$R = \prod_{i=1}^{n} r_i.$$

Функция разрушения (ненадежность системы)

$$Q = 1 - \prod_{i=1}^{n} (1 - q_i), \tag{9}$$

где $q_i = 1 - r_i$.

Так как величины $q_i \ll 1$, то в произведении, входящем в зависимость (9), можно пренебречь степенями $q_i^{\gamma} (\gamma \geqslant 2)$ и записать приближенно, что

$$Q = \sum_{i=1}^{n} q_i. \tag{10}$$

Вследствие того что функция разрушения должна иметь значение больше $Q_{\text{доп}}$, то надо найти такие рациональные пути резервирования отдельных узлов, при которых это условие будет выполнено.

Не ограничивая общности, будем предполагать, что каждый i-й узел дублируется ν_i раз.

Тогда

$$q_{ip} = q_i^{\nu_i + 1}. \tag{11}$$

В то же время масса, габариты и стоимость возрастут до величин:

$$g_{ip} = g_i(\nu_i + 1), \quad v_{ip} = v_i(\nu_i + 1), \quad c_{ip} = c_i(\nu_i + 1). \tag{12}$$

Выражая ν_i через q_i и q_{ip}, получим

$$\nu_i + 1 = \frac{x_{ip}}{x_i} \quad (e^{-x_p} = q_{ip},\ e^{-x_i} = q_i). \tag{13}$$

Следовательно,

$$g_{ip} = g_i \frac{x_{ip}}{x_i}, \quad v_{ip} = v_i \frac{x_{ip}}{x_i}, \quad c_{ip} = c_i \frac{x_{ip}}{x_i}; \tag{14}$$

общая масса системы

$$G_p = \sum_{i=1}^{n} a_i x_{ip} \leqslant G^* \quad \left(a_i = \frac{g_i}{x_i}\right), \tag{15}$$

ее общий габарит

$$V_p = \sum_{i=1}^{n} b_i x_{ip} \leqslant V^* \quad \left(b_i = \frac{v_i}{x_i}\right), \tag{16}$$

а общая стоимость

$$C_p = \sum_{i=1}^{n} d_i x_{ip} \leqslant C^* \quad \left(d_i = \frac{c_i}{x_i}\right). \tag{17}$$

Учитывая коэффициент важности β_i каждого узла, при отказе системы в целом найдем

$$Q_p = \sum_{i=1}^{n} \beta_i q_{ip} \quad \left(\sum_{i=1}^{n} \beta_i = 1\right). \tag{18}$$

Таким образом, задача проектирования системы свелась к отысканию таких $x_{ip} > 0$ в ограничениях (15) — (17), при которых величина функции разрушения (ненадежность) Q_p будет наименьшей. Рассмотрим вместо Q_p функцию $Q_p^* > Q_p$:

$$Q_p^* = \sum_{i=1}^{n} \beta_i - \sum_{i=1}^{n} \beta_i x_{ip}. \tag{19}$$

Минимум Q^* совпадает с максимумом $R_p = F$:

$$F = \sum_{i=1}^{n} \beta_i x_{ip}. \tag{20}$$

Заменив неравенства $G_p \leqslant G^*$, $V_p \leqslant V^*$, $C_p \leqslant C^*$ равенствами

$$\begin{cases} \sum\limits_{i=1}^{n} a_{ij} x_{ip} + \mu_1 = G^*, \\ \sum\limits_{i=1}^{n} b_{ij} x_{ip} + \mu_2 = V^*, \\ \sum\limits_{i=1}^{n} d_{ij} x_{ip} + \mu_3 = C^*, \end{cases} \tag{21}$$

где μ_i $(i = 1, 2, 3)$ — дополнительные неизвестные, удовлетворяющие условию $\mu_i \geqslant 0$, и задав функцию

$$F = \sum_{i=1}^{n} \beta_i x_{ip}, \tag{22}$$

получим типовую задачу линейного программирования.

Из приведенных примеров видно, что задача линейного программирования заключается в отыскании неотрицательных значений некоторых величин, обеспечивающих достижение наименьших (или наибольших) значений определенными линейными формами или величинами, зависящими от этих линейных форм.

В своей постановке задачи линейного программирования напоминают задачи вариационного исчисления (см. гл. VII). Однако в этих

задачах (в отличие от задач вариационного исчисления) речь идет об алгебраических (а не интегральных) линейных функционалах и об алгебраических конечных (но не дифференциальных или изопериметрических) условиях, которым подчинены исследуемые величины, от которых зависят функционалы.

Поэтому общая постановка задачи линейного программирования и методы решения таких задач по своему аналитическому содержанию в корне отличаются от задач и методов вариационного исчисления.

В рассмотренных примерах есть задача, содержащая линейные ограничения в виде неравенств, имеются задачи с ограничениями в виде равенств, а также задача с ограничениями обоих видов.

Задача линейного программирования с ограничениями-равенствами называется основной задачей линейного программирования.

Перейдем к общей математической формулировке основной задачи линейного программирования. Если учесть, что аналитическая постановка задачи нахождения наибольшего значения и ее решение отличаются от такой же задачи нахождения наименьшего значения только знаком, то без ограничения общности можно утверждать, что в настоящее время, в основном, все задачи линейного программирования могут найти выражение в следующей математической постановке основной задачи.

Задана система m линейных алгебраических уравнений с n неизвестными:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \quad (m < n) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \tag{23}$$

и линейная форма тех же неизвестных:

$$F = c_1x_1 + c_2x_2 + \cdots + c_nx_n + c_0. \tag{24}$$

Найти такие неотрицательные значения неизвестных

$$x_i \geqslant 0 \quad (i = 1, 2, 3, \ldots, n), \tag{25}$$

при которых линейная форма F принимает наименьшее значение (минимизируется).

В различных условиях целесообразно использовать различные виды записи основной задачи. Наиболее часто встречаются следующие виды [1]:

а) Минимизировать линейную форму

$$F = c_0 + \sum_{i=1}^{n} c_i x_i \tag{26}$$

при условиях

$$x_i \geqslant 0 \quad (i = 1, 2, 3, \ldots, n)$$

[1] См.: С. Гасс. Линейное программирование. М., Физматгиз, 1961, с. 56.

и

$$\sum_{k=1}^{n} a_{jk}x_k = b_j \quad (j = 1, 2, 3, \ldots, m), \quad (m < n).$$

Этот вид использован в приведенной формулировке задачи. Не изменяя существа задачи, можно, очевидно, считать $c_0 = 0$.

б) Минимизировать форму $c_0 + \boldsymbol{cx}$ при условии

$$\boldsymbol{x} \geqslant \boldsymbol{0}, \quad A\boldsymbol{x} = \boldsymbol{b},$$

где $\boldsymbol{c} = \|c_1 c_2 \ \ldots \ c_n\|$ — вектор-строка,

$$\boldsymbol{x} = \begin{Vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Vmatrix} \text{— вектор-столбец,}$$

$$A = \begin{Vmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{Vmatrix} \quad (m < n),$$

$$\boldsymbol{b} = \begin{Vmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{Vmatrix}.$$

в) Минимизировать форму

$$c_0 + \boldsymbol{cx}$$

при условии $\boldsymbol{x} \geqslant \boldsymbol{0}$

и

$$\sum_{i=1}^{n} x_i \boldsymbol{P}_i = \boldsymbol{P}_0,$$

где $\boldsymbol{P}_i$ — i-й столбец матрицы A ($i = 1, 2, 3, \ldots, n$);

$$\boldsymbol{P}_0 = \boldsymbol{b}.$$

Условимся называть допустимым решением основной задачи линейного программирования любую совокупность переменных

$$x_1 \geqslant 0, \quad x_2 \geqslant 0, \quad \ldots, \quad x_n \geqslant 0,$$

удовлетворяющую уравнениям (23).

Оптимальным решением будем называть то из допустимых решений, при котором линейная форма (24) принимает наименьшее значение (минимизируется).

Основная задача линейного программирования не обязательно имеет решение. Может оказаться, что уравнения (23) противоречат друг другу или что они имеют решения, но не в области неотрицательных значений $x_1, x_2, \ldots, x_n$. Тогда основная задача линейного

программирования не имеет допустимых решений. Наконец, может случиться, что допустимые решения основной задачи существуют, но среди них нет оптимального: функция F в области допустимых решений не ограничена снизу.

Вопрос о существовании допустимых решений основной задачи линейного программирования рассматривается с помощью математического аппарата линейной алгебры (см. гл. VI).

В линейной алгебре доказывается, что *для совместности системы линейных уравнений* (23), *необходимо и достаточно, чтобы ранг матрицы системы был равен рангу ее расширенной матрицы.* Этот общий ранг r называется р а н г о м с и с т е м ы. Он представляет собой не что иное, как число линейно-независимых уравнений среди наложенных в задаче ограничений-равенств.

Очевидно, ранг системы r не может быть больше числа уравнений m и также не может быть больше общего числа переменных n.

Структура задачи линейного программирования существенно зависит от ранга системы уравнений (23).

Рассмотрим, прежде всего, случай, когда $r = n$, т. е. когда число линейно-независимых уравнений, входящих в систему (23), равно числу переменных n. Из линейной алгебры известно, что в таком случае система уравнений (23) имеет единственное решение: x_1, x_2, ..., x_n. Если в этом решении хотя бы одна из величин x_1, x_2, ..., x_n отрицательна, то это означает, что полученное решение недопустимо и, значит, основная задача линейного программирования не имеет решения. Если все величины $x_1, x_2, \ldots, x_n$ неотрицательны, то найденное решение является допустимым. Оно же, очевидно, и оптимальное, так как других решений нет.

В случае, когда $r < n$ (т. е. когда число линейно-независимых уравнений, которым должны удовлетворять переменные x_1, x_2, ..., x_n, меньше числа самих переменных), система (23) либо не имеет решений, либо имеет бесчисленное множество решений. Если среди этого бесчисленного множества решений нет ни одного решения, для которого все $x_1, x_2, \ldots, x_n$ неотрицательны, то это означает, что основная задача линейного программирования не имеет допустимого решения. Если же среди множества решений имеются такие, для которых все $x_1, x_2, \ldots, x_n$ неотрицательны, то каждое из них допустимо. Здесь возникает задача: найти среди всех допустимых решений оптимальное, т. е. такое решение $x_i \geqslant 0$ $(i = 1, 2, 3, \ldots, n)$, для которого линейная функция (форма) $F = \sum_{i=1}^{n} c_i x_i$ обращается в минимум (минимизируется).

Из приведенных в начале параграфа примеров видно, что в задачах линейного программирования ограничения, наложенные на переменные $x_1, x_2, \ldots, x_n$, могут иметь вид линейных неравенств. При этом может случиться, что в некоторых из них стоит знак $\geqslant$, а в других $\leqslant$. В таком случае желательно все ограничения-неравенства записать в единой форме, например в виде $\geqslant$. Для этого следует в неравенствах вида $\leqslant$ лишь переменить знаки обеих частей. Тогда

получим задачу линейного программирования с ограничениями-неравенствами:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + b_1' \geqslant 0, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + b_2' \geqslant 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n + b_m' \geqslant 0, \end{cases} \quad \begin{pmatrix} b_i' = -b_i, \\ i = 1, 2, \ldots, m \end{pmatrix} \quad (27)$$

где требуется найти такую совокупность неотрицательных значений $x_1, x_2, \ldots, x_n$, которая удовлетворяла бы системе неравенств (27) и, кроме того, обращала бы в минимум линейную функцию тех же переменных $F = c_1x_1 + c_2x_2 + \ldots + c_nx_n$.

От полученной задачи с ограничениями-неравенствами легко перейти к основной задаче линейного программирования[1]. Действительно, введем обозначения

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + b_1', \\ y_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + b_2', \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_m = a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n + b_m', \end{cases} \quad (28)$$

где $y_1, y_2, \ldots, y_m$ — некоторые новые переменные, которые мы будем называть «добавочными» (б а з и с н ы м и). Согласно условиям (27), эти добавочные переменные так же, как и «первоначальные» (свободные) $x_1, x_2, \ldots, x_n$, должны быть неотрицательными.

Таким образом, здесь возникает задача линейного программирования в следующей постановке: найти неотрицательные значения $n + m$ переменных $x_1, x_2, \ldots, x_n; y_1, y_2, \ldots, y_m$, которые удовлетворяли бы системе уравнений (28) и одновременно обращали в минимум функцию $F = c_1x_1 + c_2x_2 + \ldots + c_nx_n$.

Как видно, перед нами в чистом виде основная задача линейного программирования. Уравнения (23) заданы в форме, уже разрешенной относительно базисных переменных $y_1, y_2, \ldots, y_m$ которые выражены через свободные переменные $x_1, x_2, \ldots, x_n$. Общее количество переменных равно $n + m$, из них n «первоначальных» и m «добавочных». Функция F выражена только через «первоначальные» переменные (коэффициенты при «добавочных» переменных в ней равны нулю).

Итак, задача линейного программирования с ограничениями-неравенствами сведена к основной задаче линейного программирования, но с большим, чем было, числом переменных.

Всегда возможен и обратный переход — от основной задачи линейного программирования к эквивалентной ей (имеющей то же решение) задаче с ограничениями-неравенствами. И если при прямом переходе мы увеличивали число переменных, то при обратном пере-

[1] См.: Е. С. В е н т ц е л ь. Исследование операций. М., Советское радио, 1972, с. 55—57.

ходе будем его уменьшать, устраняя базисные переменные и оставляя только свободные.

Если среди ограничений задачи есть как уравнения, так и неравенства, то рекомендуется произвести унификацию и перейти к единообразной форме, например, к форме основной задачи линейного программирования.

Покажем на задаче «Об использовании сырья» (см. § 44), как осуществляется переход от общей математической постановки задачи к основной задаче линейного программирования, а затем к ее форме с ограничениями-неравенствами.

Систему неравенств (4) перепишем в виде:

$$\begin{cases} 0 \leqslant 19 - 2x_1 - 3x_2, \\ 0 \leqslant 13 - 2x_1 - x_2, \\ 0 \leqslant 15 - 3x_2, \\ 0 \leqslant 18 - 3x_1. \end{cases} \tag{29}$$

Введем четыре новых («добавочных») неизвестных:

$$\begin{cases} x_3 = 19 - 2x_1 - 3x_2, \\ x_4 = 13 - 2x_1 - x_2, \\ x_5 = 15 - 3x_2, \\ x_6 = 18 - 3x_1 \end{cases} \tag{30}$$

и вместо формы F (5) — форму

$$F_1 = -F = -7x_1 - 5x_2. \tag{31}$$

Теперь задача приобрела вид основной задачи линейного программирования: найти неотрицательное решение системы (30), при котором форма F_1 принимает наименьшее значение.

Далее упорядочим систему (30)

$$\begin{cases} 2x_1 + 3x_2 + x_3 = 19, \\ 2x_1 + x_2 + x_4 = 13, \\ 3x_2 + x_5 = 15, \\ 3x_1 + x_6 = 18 \end{cases}$$

и выпишем для нее матрицу A:

$$A = \begin{Vmatrix} 2 & 3 & 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 & 1 & 0 \\ 3 & 0 & 0 & 0 & 0 & 1 \end{Vmatrix}.$$

Ранг этой матрицы равен 4 ($r = 4$), количество неизвестных: $n + m = 6$. Поэтому $n + m - r = 2$ — имеются две свободные неизвестные. Четыре неизвестные надо выразить через две. Здесь в качестве свободных неизвестных удобно взять x_1 и x_2. Выше уже x_3, x_4, x_5 и x_6 и форма F_1 выражены через x_1 и x_2.

Тогда задача линейного программирования выражается в искомой форме ограничений-неравенств следующим образом: необходимо найти такое решение системы неравенств:

$$\begin{cases} x_1 \geqslant 0, \\ x_2 \geqslant 0, \\ 19 - 2x_1 - 3x_2 \geqslant 0, \\ 13 - 2x_1 - x_2 \geqslant 0, \\ 15 - 3x_2 \geqslant 0, \\ 18 - 3x_1 \geqslant 0, \end{cases} \tag{32}$$

при котором линейная форма $F_1 = -7x_1 - 5x_2$ принимает наименьшее значение.

Всякая задача линейного программирования всегда связана с некоторыми целевыми установками, определяющими выбор ее решения. Целевая установка задачи задается линейной формой F, которая называется критерием (показателем) качества (в настоящее время эта линейная форма чаще всего называется функцией цели или целевой функцией).

Решение задач линейного программирования с числом свободных переменных, существенно большим двух-трех, путем простого перебора огромного количества различных вариантов не может быть осуществлено даже с использованием самой совершенной вычислительной техники. Поэтому общей теорией линейного программирования в настоящее время создано немало удобных для практики методов решения оптимизационных задач. Одним из наиболее эффективных является симплексный метод (метод последовательного улучшения плана). Он применим ко всем линейным оптимизационным задачам и позволяет получить точное численное решение задачи, но при большом числе переменных становится настолько громоздким, что для выполнения вычислительной работы требует использования ЭВМ. Ниже, в § 46, будут изложены основные положения этого метода.

§ 45. ГЕОМЕТРИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ЗАДАЧИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ И ЕЕ РЕШЕНИЯ

Для лучшего понимания особенностей решения задачи линейного программирования удобно воспользоваться ее геометрическим истолкованием. При этом следует иметь в виду, что геометрический подход оказывается малополезным для получения численных решений большинства практических задач линейного программирования[1].

Предварительно рассмотрим некоторые понятия аналитической геометрии в n-мерном векторном пространстве, не прибегая к доказательствам приводимых утверждений.

[1] См.: Г. Вагнер. Основы исследования операций. М., Мир, 1972, т. I, с. 110.

Некоторые понятия аналитической геометрии в n-мерном пространстве

1. Пусть в n-мерном линейном векторном пространстве R_n задан вектор $x = (x_1, x_2, \ldots, x_n)$. Если координаты некоторой точки M совпадают численно с координатами вектора x, то вектор x называется радиусом-вектором точки M. Всякий упорядоченный набор n чисел определяет некоторый вектор, а вместе с ним, как с радиусом-вектором, и некоторую точку с n-координатами. Совокупность всех таких точек образует n-мерное точечное пространство, которое будем обозначать через T_n (см. § 34). Различие между пространствами R_n и T_n состоит в различии объектов, образующих эти пространства (элементами пространства R_n являются векторы, элементами пространства T_n — точки).

2. Точка 0 с координатами (0; 0; 0; ...; 0) называется началом координат, соответствующий ей радиус-вектор — нулевым вектором (нуль-вектором). Множество точек, у которых все координаты, кроме одной x_k, нули, а координата x_k произвольна, называется координатной осью Ox_k (в n-мерном пространстве T_n, очевидно, имеется n координатных осей Ox_1, Ox_2, ..., Ox_n).

3. Множество точек, у которых одна координата $x_k = 0$, а все остальные отличны от нуля, т. е. совокупность точек вида $(x_1, x_2, \ldots, x_{k-1}, 0, x_{k+1}, \ldots, x_n)$ называется координатной гиперплоскостью x_k (в n-мерном пространстве T_n имеется n координатных гиперплоскостей $x_1 = 0$, $x_2 = 0$, ..., $x_n = 0$).

Вообще же *гиперплоскостью в n-мерном пространстве T_n называется множество точек, удовлетворяющих уравнению*

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n + a_0 = 0,$$

где a_i ($i = 0, 1, 2, \ldots, n$) — произвольные, постоянные для данной гиперплоскости, действительные числа.

Если $a_0 = 0$, то гиперплоскость проходит через начало координат, если $a_k = 0$, то гиперплоскость не пересекает координатной оси Ox_k (проходит параллельно ей).

Множество точек, принадлежащих одновременно двум гиперплоскостям

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n + a_0 = 0 \tag{33}$$

и

$$b_1x_1 + b_2x_2 + \cdots + b_nx_n + b_0 = 0, \tag{34}$$

называется их пересечением.

Пересечение гиперплоскостей записывается системой уравнений этих гиперплоскостей:

$$\begin{cases} a_1x_1 + a_2x_2 + \cdots + a_nx_n + a_0 = 0, \\ b_1x_1 + b_2x_2 + \cdots + b_nx_n + b_0 = 0. \end{cases} \tag{35}$$

Необходимым и достаточным условием отсутствия пересечения двух гиперплоскостей (33) и (34) является условие

$$\frac{a_1}{b_1} = \frac{a_2}{b_2} = \ldots = \frac{a_n}{b_n} \neq \frac{a_0}{b_0}. \qquad (36)$$

Уравнения (33) и (34) определяют одну и ту же гиперплоскость, если

$$\frac{a_1}{b_1} = \frac{a_2}{b_2} = \ldots = \frac{a_n}{b_n} = \frac{a_0}{b_0}. \qquad (37)$$

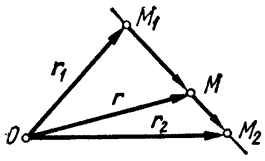

Рис. 95

Две гиперплоскости называются п а р а л л е л ь н ы м и, если они не пересекаются или совпадают, т. е. при выполнении условий (36) или (37).

4. Если двум точкам M_1 и M_2 соответствуют радиусы-векторы $\overrightarrow{OM_1} = \boldsymbol{r}_1$ и $\overrightarrow{OM_2} = \boldsymbol{r}_2$ ($\overrightarrow{M_1M_2} = \boldsymbol{r}_2 - \boldsymbol{r}_1$), то всякий вектор $\boldsymbol{\rho} = \overrightarrow{\lambda M_1M_2}$ ($\lambda = \text{const}$) коллинеарен вектору $\overrightarrow{M_1M_2}$, а о т р е з к о м $\overrightarrow{M_1M_2}$ называется совокупность точек, радиусы-векторы которых удовлетворяют равенству $\boldsymbol{r} = \boldsymbol{r}_1 + \theta(\boldsymbol{r}_2 - \boldsymbol{r}_1)$ или, что то же,— равенству

$$\boldsymbol{r} = (1 - \theta)\boldsymbol{r}_1 + \theta\boldsymbol{r}_2, \qquad (38)$$

где $0 \leqslant \theta \leqslant 1$. Поэтому в соответствии с рис. 95 можно записать: $\overrightarrow{M_1M} \parallel \overrightarrow{M_1M_2}$, $\boldsymbol{r} = \boldsymbol{r}_1 + \overrightarrow{M_1M}$, $\overrightarrow{M_1M_2} = \boldsymbol{r}_2 - \boldsymbol{r}_1$, $\overrightarrow{M_1M} = \overrightarrow{M_1M_2} \cdot \theta$.

Основываясь на свойствах линейных операций над векторами (см. гл. VI) и считая

$$\boldsymbol{r}_1 = (x_{11}, x_{12}, \ldots, x_{1n}), \quad \boldsymbol{r}_2 = (x_{21}, x_{22}, \ldots, x_{2n}),$$
$$\boldsymbol{r} = (x_1, x_2, \ldots, x_n),$$

можем записать для любой точки отрезка M_1M_2:

$$x_i = (1 - \theta)x_{1i} + \theta x_{2i} \quad (i = 1, 2, \ldots, n; \quad 0 \leqslant \theta \leqslant 1). \qquad (39)$$

Точка, принадлежащая одновременно отрезку M_1M_2, для которого $\boldsymbol{r} = (1 - \theta)\boldsymbol{r}_1 + \theta\boldsymbol{r}_2$ ($0 \leqslant \theta \leqslant 1$), и гиперплоскости П, для которой $a_1x_1 + a_2x_2 + \ldots + a_nx_n + a_0 = 0$, называется т о ч к о й п е р е с е ч е н и я отрезка M_1M_2 и гиперплоскости П.

Все пространство T_n с помощью произвольной гиперплоскости П разбивается на две части так, что выполняются условия: любой отрезок, соединяющий две точки из одной какой-либо части, не пересекает гиперплоскости П, а отрезок, соединяющий пару точек из различных частей, пересекает гиперплоскость П. Каждая из таких частей пространства T_n называется п о л у п р о с т р а н с т в о м. Сама гиперплоскость П может быть отнесена как к одному, так и к другому полупространству.

5. *Совокупность V точек n-мерного пространства T_n называется выпуклым телом, если любой отрезок M_1M_2, концевые точки которого принадлежат множеству V, целиком принадлежит этому множеству.*

Простейшим выпуклым телом является гиперплоскость. Выпуклым телом оказывается также полупространство.

Пересечением тел называется совокупность точек, принадлежащих каждому из этих тел. Пересечение выпуклых тел является также выпуклым телом. Пересечение любого количества полупространств, очевидно, является выпуклым телом.

6. Пусть дана линейная форма

$$F = a_1x_1 + a_2x_2 + \cdots + a_nx_n + a_0, \tag{40}$$

и пусть точка M_0 имеет координаты $(x_{10}, x_{20}, \ldots, x_{n0})$.

Число $F_0 = a_1x_{10} + a_2x_{20} + \ldots + a_nx_{n0} + a_0$ называется значением линейной формы F в точке M_0 и обозначается $F(M_0)$.

Придавая координатам точки M_0 различные значения, можно получать соответствующие различные значения линейной формы F. Нетрудно видеть, что множеством точек $M(x_1, x_2, \ldots, x_n)$, в которых форма F имеет одно и то же значение c, оказывается гиперплоскостью. Действительно, если $F(M) = c$, то

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n + a_0 = c \tag{41}$$

или

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n + (a_0 - c) = 0,$$

а уравнение (41) определяет гиперплоскость. Такая гиперплоскость называется гиперплоскостью равных значений линейной формы F.

Две гиперплоскости равных значений линейной формы

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n + a_0 = c_1$$

и

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n + a_0 = c_2,$$

соответствующие различным значениям c ($c_1 \neq c_2$), не пересекаются и поэтому параллельны между собой.

7. Рассмотрим линейное неравенство

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n + a_0 \geqslant 0. \tag{42}$$

Совокупность всех точек пространства T_n, удовлетворяющих своими координатами неравенству (42), называется областью решений этого неравенства.

Оказывается (и это легко проверить), что областью решений линейного неравенства является полупространство. Из этого следует, что областью решений системы линейных неравенств является некоторое выпуклое тело, образованное пересечением выпуклых тел — полупространств. Такое выпуклое тело называется выпуклым многогранником.

Итак, *областью решений системы линейных неравенств является выпуклый многогранник, ограниченный гиперплоскостями, уравнения которых получаются из линейных неравенств заданной системы путем замены знаков неравенств на знаки точного равенства.*

Могут встретиться три различных вида областей решений системы неравенств. Покажем эти виды на примерах из пространства T_2:

а) Область решений системы неравенств

$$\begin{cases} 2x_1 - 3x_2 + 13 \geqslant 0, & \text{(I)} \\ x_1 + x_2 - 6 \geqslant 0, & \text{(II)} \\ 4x_1 - x_2 - 19 \leqslant 0, & \text{(III)} \end{cases}$$

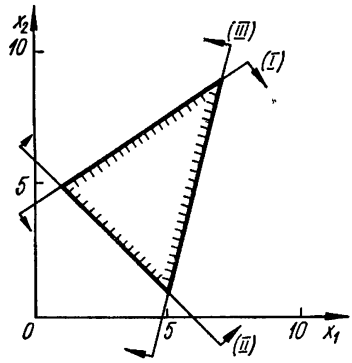

Рис. 96

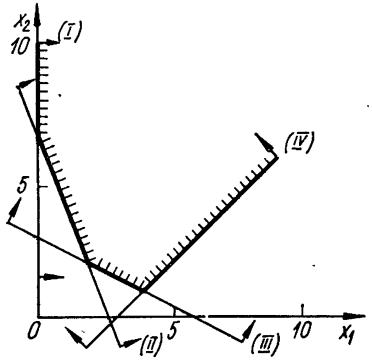

Рис. 97

представляет собой ограниченное выпуклое тело — треугольник, изображенный на рис. 96. (Здесь и дальше стрелками указаны полуплоскости, являющиеся областями решений соответствующих неравенств системы). Система неравенств определенная.

б) Область решений системы неравенств

$$\begin{cases} x_1 \geqslant 0, & \text{(I)} \\ 3x_1 + x_2 - 8 \geqslant 0, & \text{(II)} \\ x_1 + 2x_2 - 6 \geqslant 0, & \text{(III)} \\ x_1 - x_2 - 3 \leqslant 0 & \text{(IV)} \end{cases}$$

является неограниченным выпуклым телом (рис. 97). Система неравенств неопределенная.

в) Область решений системы неравенств

$$\begin{cases} 2x_1 - 3x_2 + 13 \leqslant 0, & \text{(I)} \\ x_1 + x_2 - 6 \leqslant 0, & \text{(II)} \\ 4x_1 - x_2 - 19 \geqslant 0 & \text{(III)} \end{cases}$$

пустая — не содержит ни одной точки. Система неравенств несовместна (рис. 98).

Рассмотренные положения аналитической геометрии в n-мерном пространстве позволяют теперь перейти к дальнейшему изучению задачи линейного программирования.

Геометрическое представление решения задачи линейного программирования. Рассмотрим случай, когда число переменных n на два больше, чем число независимых уравнений m, которым они должны удовлетворять: $n - m = 2$ (двумерная задача).

В этом параграфе для простоты будем считать уравнения задачи программирования линейно-независимыми: при этом ранг системы r будет равен числу уравнений m.

В этом случае число свободных переменных равно двум, и решение задачи линейного программирования может получить наглядное геометрическое истолкование на плоскости (в пространстве T_2), которое поможет глубже понять применяемые аналитические методы.

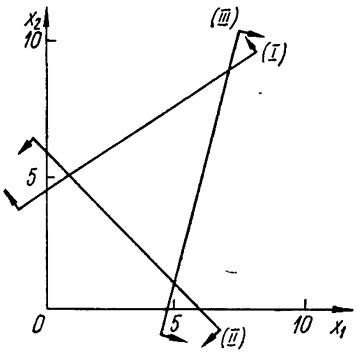

Рис. 98

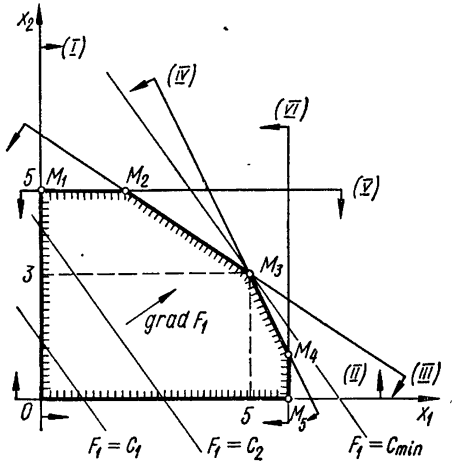

Рис. 99

Для выявления особенностей решения воспользуемся примером задачи «Об использовании сырья», которая приведена выше к форме с ограничениями-неравенствами (32) по минимизации целевой функции $F_1 = -7x_1 - 5x_2$. В этой задаче имеются две свободные переменные x_1 и x_2.

Введем прямоугольную систему координат x_1Ox_2 и построим выпуклый многоугольник (в общем случае выпуклый многогранник) соответствующий ограничениям-неравенствам системы (32).

$$\begin{cases} x_1 \geqslant 0, & \text{(I)} \\ x_2 \geqslant 0, & \text{(II)} \\ 19 - 2x_1 - 3x_2 \geqslant 0, & \text{(III)} \\ 13 - 2x_1 - x_2 \geqslant 0, & \text{(IV)} \\ 15 - 3x_2 \geqslant 0, & \text{(V)} \\ 18 - 3x_1 \geqslant 0. & \text{(VI)} \end{cases}$$

Наряду с многоугольником (рис. 99) рассмотрим линейную форму $F_1 = -7x_1 - 5x_2$.

Эта форма — линейная функция координат x_1 и x_2 некоторой точки на плоскости.

Так как нас интересует значение формы F_1, то необходимо построить прямые (в общем случае — гиперплоскости) равных значений F_1, т. е. прямые вида

$$-7x_1 - 5x_2 = C \quad (C = \text{const}). \tag{43}$$

Меняя С, будем получать различные, но параллельные между собой прямые, соответствующие фиксированным значениям C (C_1, C_2, ..., C_n).

Рассматривая F_1 как функцию, задающую скалярное поле, установим, что уравнение $F_1 = C$ определяет линии уровня (в общем случае гиперплоскости уровня)[1].

Нас интересует такое положение линии уровня, при котором она пересекает многоугольник допустимых значений (проходит через точки области существования решений).

При этом, так как форма F_1 — линейная функция, то совершенно очевидно, что свое наименьшее значение она может достигнуть только на границе области.

В рассматриваемой задаче многоугольник допустимых значений замкнутый, линия уровня $-7x_1 - 5x_2 = C$ не параллельна ни одной стороне многоугольника, а потому задача имеет решение и притом единственное. Найдем

$$\operatorname{grad} F_1 = \left\{-\frac{\partial F_1}{\partial x_1}\ ;\ -\frac{\partial F_1}{\partial x_2}\right\} = \{7;\ 5\}.$$

Придадим постоянной C какое-либо значение, но такое, чтобы соотвествующая ему линия уровня пересекала область решений, а затем, перемещая эту прямую параллельно самой себе (меняя значение C) в направлении градиента, увидим, что наступит момент, когда прямая выйдет за пределы многоугольника. Это произойдет в точке M_3 (5; 3). Этой точкой и определяется оптимальное решение: $x_1 = 5$, $x_2 = 3$. Наименьшее значение линейной формы F_1 будет

$$F_{1\,\min} = -7 \cdot 5 - 5 \cdot 3 = -50.$$

Таким образом, в соответствии с требованием задачи для наиболее рационального плана использования сырья, гарантирующего предприятию наибольший доход, следует выпускать 5 единиц продукции вида Π_1 и 3 единицы продукции вида Π_2. Максимальный доход будет: $F_{\max} = 50$ рублей.

При решении двумерной задачи линейного программирования описанным выше способом могут иметь место следующие случаи:

а) Многоугольник допустимых значений замкнутый, и $F_{\min}$ достигается в одной вершине многоугольника (рис. 100, *а*), как это было в рассмотренной выше задаче. В этом случае имеется единственное решение (оно же и оптимальное).

[1] Эти термины применяются в линейном программировании.

б) Многоугольник допустимых значений замкнутый, и F_{min} достигается во всех точках одной из его сторон (рис. 100, *б*). Здесь задача имеет бесчисленное множество оптимальных решений.

в) Многоугольник допустимых значений не замкнутый, и линия уровня $F = C$ пересекает его границу только с одной стороны (рис. 100, *в*). В этом случае задача не имеет оптимального решения, так как в области допустимых решений линейная функция F не ограничена снизу — в задаче не хватает условий.

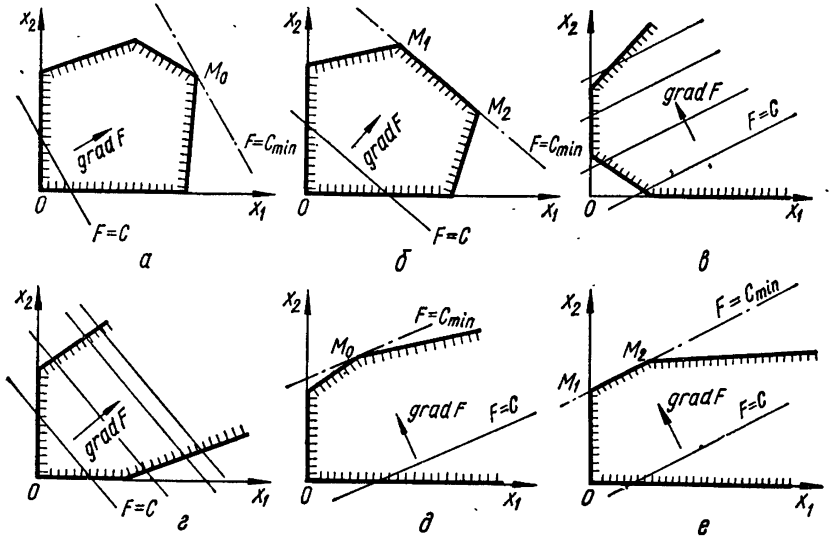

Рис. 100

г) Многоугольник допустимых значений не замкнутый, и линия уровня вообще не выходит за пределы многоугольника (рис. 100, *г*). Здесь задача не имеет оптимального решения, так как линейная форма F нигде не достигает наименьшего значения.

д) Многоугольник допустимых значений не замкнутый и F_{min} достигается в вершине многоугольника (рис. 100, *д*). В этом случае задача имеет единственное оптимальное решение.

е) Многоугольник допустимых значений не замкнутый, и F_{min} достигается во всех точках одной из его сторон (рис. 100, *е*). Здесь, как и в случае б), задача имеет бесчисленное множество оптимальных решений (решение неопределенно).

Решение, лежащее в одной из вершин многоугольника допустимых значений, называется опорным решением, а сама величина — опорной точкой.

Для того чтобы найти оптимальное решение, в принципе достаточно перебрать все вершины многоугольника допустимых значений (а не рассматривать множество «слепых» вариантов) и выбрать из небольшого числа опорных точек ту, где целевая функция F достигает минимума.

Если число свободных переменных в задаче линейного программирования равно 2, а число базисных переменных m, и решение задачи существует, то оно всегда достигается в точке, где, по крайней мере, две из переменных $x_1, x_2, \ldots, x_{m+1}, x_{m+2}$ обращаются в нуль. Действительно, в любой опорной точке пересекаются по крайней мере две из ограничивающих прямых; могут же в ней пересекаться и более двух прямых.

Для трехмерной задачи линейного программирования, когда число переменных n на 3 превышает число независимых уравнений m ($n - m = 3$, т. е. число свободных переменных равно 3), геометрическое представление решения строится уже не на плоскости (как это было сделано выше для двумерной задачи), а в трехмерном эвклидовом пространстве T_3. Здесь каждое условие $x_i = 0$ ($i = 1,2, \ldots, n - 3$) для одной из базисных переменных геометрически изображается уже не прямой, а плоскостью. По одну сторону от этой плоскости $x_i > 0$, по другую $x_i < 0$. Координатные плоскости x_2Ox_3, x_1Ox_3 и x_1Ox_2 изображают условия $x_1 = 0$, $x_2 = 0$, $x_3 = 0$ соответственно. Область допустимых решений (если она существует) представляет собой выпуклый многогранник, ограниченный этими плоскостями, т. е. часть пространства, для которой выполнены все условия:

$$x_1 \geqslant 0, \quad x_2 \geqslant 0, \quad x_3 \geqslant 0, \ldots, x_n \geqslant 0.$$

Роль линии уровня в этом случае будет играть гиперплоскость уровня, уравнение которой $F_{\text{ур}} = 0$, где

$$F_{\text{ур}} = F - \gamma_0, \quad F = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3.$$

При перемещении этой плоскости параллельно самой себе в одну сторону $F_{\text{ур}}$ будет убывать, в другую — возрастать. Точка A, в которой достигается оптимальное решение (если оно существует), представляет собой ту вершину многогранника допустимых значений, которая находится дальше всего от начала координат, считая по направлению убывания $F_{\text{ур}}$. Может оказаться, как при $n - m = 2$, что задача линейного программирования имеет бесчисленное множество решений, заполняющих либо целое ребро, либо целую грань многогранника допустимых значений (решений). Оптимальное решение x_1^*, x_2^*, x_3^* (если оно существует) совпадает с одной из опорных точек, т. е. вершин многогранника, в которой по крайней мере три из переменных $x_1, x_2, \ldots, x_n$ обращаются в нуль.

Пользоваться геометрическим представлением для непосредственного отыскания решения задачи линейного программирования даже при $n - m = 3$ затруднительно; при $n - m = k > 3$ это, вообще, выводит нас за рамки трехмерного пространства, и «геометрическое представление» теряет наглядность. Однако соответствующие понятия аналитической геометрии k-мерного векторного пространства оказываются удобными. Здесь можно говорить об области допустимых решений как о некотором выпуклом многограннике («сверхмногограннике») в пространстве k измерений, ограни-

ченном m гиперплоскостями; об оптимальном решении как одной из «вершин» этого многогранника; о каждой «вершине» как об «опорной точке» и т. д. Такими геометрическими терминами можно, по желанию, пользоваться или не пользоваться.

Основное назначение геометрической интерпретации состоит в том, что она позволяет обосновать и сформулировать следующие свойства основной задачи линейного программирования при любых значениях числа переменных n и числа уравнений m ($m < n$):

1. Оптимальное решение, если оно существует, лежит не внутри, а на границе области допустимых решений, в одной из опорных точек, в каждой из которых по крайней мере k из переменных обращаются в нуль. (Случай, когда в оптимальном решении, при числе свободных переменных, равном k, обращаются в нуль не k, а больше переменных, называется в ы р о ж д е н н ы м).

2. Для того чтобы найти оптимальное решение, нужно, переходя от одной опорной точки к другой, двигаться в направлении уменьшения линейной функции F, которую требуется минимизировать.

На этих принципах и основаны вычислительные методы решения основной задачи линейного программирования, главным из которых является симплексный метод.

§ 46. СИМПЛЕКСНЫЙ МЕТОД

Перед непосредственным изложением симплексного метода в общем виде рассмотрим ряд необходимых определений и продемонстрируем идею этого метода на конкретном примере решения основной задачи линейного программирования.

а) Планом задачи линейного программирования называется вектор

$$x = \begin{Vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Vmatrix},$$

удовлетворяющий условиям

$$x_i \geqslant 0 \quad (i = 1, 2, \ldots, n)$$

и

$$\sum_{k=1}^{n} a_{jk}x_k = b_j \quad (j = 1, 2, \ldots, m), \quad (m < n).$$

б) План x называется опорным планом, если векторы P_i, входящие в разложение $P_0 = \sum_{i=1}^{n} x_i P_i$ с положительными коэффициентами, линейно-независимы (см. третий вид записи основной задачи, с. 300).

Непосредственно из этого определения следует, что число положительных составляющих опорного плана не может быть больше, чем m (так как матрица A имеет всего m строк, а $m < n$).

в) Опорный план называется невырожденным, если он содержит точно m положительных составляющих.

г) Оптимальным планом (или решением) задачи линейного программирования называется план, минимизирующий линейную форму

$$F = c_0 + \sum_{i=1}^{n} c_i x_i.$$

д) Функция $f(x)$, определенная в n-мерном пространстве и принимающая действительные значения, называется линейным функционалом, если для любых векторов x и y из этого пространства и чисел α и β выполняется равенство $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$.

Очевидно, линейная форма является линейным функционалом.

Далее не представляет большого труда установить для основной задачи линейного программирования следующие важные факты:

1. Множество K всех планов выпукло.

2. Каждый опорный план соответствует крайней точке множества K.

3. С каждой крайней точкой множества K связаны m линейно-независимых векторов из данной системы $P_1, P_2, \ldots, P_n$ n векторов.

4. Линейная форма F задачи достигает своего минимума в крайней точке выпуклого множества K всех планов; если минимум достигается более чем в одной точке, то он достигается и в любой выпуклой линейной комбинации этих точек (т. е. на отрезке, соединяющем две крайние точки).

Таким образом, отыскание минимума линейной формы задачи всегда связано с отысканием опорных планов — только опорный план (точнее один из опорных планов) может быть решением основной задачи линейного программирования.

Симплекс-метод, о котором пойдет речь ниже, позволяет, зная некоторый опорный план задачи, путем конечного числа шагов получить оптимальный план.

Каждый шаг представляет собой переход от одного опорного плана к другому — переход от одной крайней точки выпуклого множества планов к другой — с последующим уменьшением значений минимизируемой формы F. Вследствие этого симплексный метод получил еще название метода последовательного улучшения планов.

Покажем основную идею симплексного метода и его применение на примере решения задачи «Об использовании сырья».

Система ограничений-равенств для этой задачи имеет вид (30):

$$\begin{cases} x_3 = 19 - 2x_1 - 3x_2, \\ x_4 = 13 - 2x_1 - x_2, \\ x_5 = 15 - 3x_2, \\ x_6 = 18 - 3x_1, \end{cases}$$

а линейная форма (31): $F_1 = -7x_1 - 5x_2$.

Система (30) совместна, количество неизвестных $n = 6$, ранг системы $r = 4$, количество свободных неизвестных $n - r = k = 2$; в качестве свободных неизвестных выберем x_1 и x_2.

Так как все $x_i \geqslant 0$ $(i = 1, 2, \ldots, 6)$, то наименьшие значения свободных неизвестных будут $x_1 = 0$ и $x_2 = 0$. При этом базисные (не свободные) неизвестные имеют значения $x_3 = 19$, $x_4 = 13$, $x_5 = 15$, $x_6 = 18$.

Эти значения положительны, допустимы, а следовательно, вектор

$$x = \begin{Vmatrix} 0 \\ 0 \\ 19 \\ 13 \\ 15 \\ 18 \end{Vmatrix} \tag{44}$$

представляет собой один из возможных планов.

При этом минимизируемая форма

$$F_1(x) = 0.$$

Так как обе свободные неизвестные входят в состав линейной формы F_1 с отрицательными коэффициентами, то увеличение x_1 и x_2 повлечет за собой уменьшение формы. Однако неограниченное увеличение x_1 и x_2 приведет к появлению отрицательных значений базисных неизвестных, что недопустимо. Поэтому увеличивать значения x_1 и x_2 надо, сообразуясь с изменением значений остальных неизвестных.

Отметим, что из третьего уравнения системы ограничений-равенств (30) видно, что если $x_2 > 5$, то x_5 будет отрицательным. При $x_2 = 5$ будет $x_5 = 0$. Сохраним $x_1 = 0$, возьмем $x_5 = 0$, тогда $x_2 = 5$, $x_3 = 4$, $x_4 = 8$, $x_6 = 18$ — все значения неизвестных неотрицательны. Поэтому

$$x = \begin{Vmatrix} 0 \\ 5 \\ 4 \\ 8 \\ 0 \\ 18 \end{Vmatrix} \tag{45}$$

также будет планом.

При этом замечаем, что при новом плане значение минимизируемой линейной формы $F_1(x) = -25$ меньше, чем при первоначальном плане.

В каждом из двух планов (44) и (45) по два неизвестных имеют нулевые значения. Но во втором случае ($x_1 = 0$, $x_5 = 0$) значение формы F_1 оказалось меньше. Быть может в качестве свободных неизвестных выбрать x_1 и x_5?

Сделаем это; тогда система ограничений-равенств примет вид

$$\begin{cases} x_2 = 5 - \frac{1}{3} x_5, \\ x_3 = 4 - 2x_1 + x_5, \\ x_4 = 8 - 2x_1 + \frac{1}{3} x_5, \\ x_6 = 18 - 3x_1, \end{cases} \tag{46}$$

а линейная форма

$$F_1 = -25 - 7x_1 + \frac{5}{3} x_5. \tag{47}$$

Теперь x_1 и x_5 не могут быть взяты меньше нуля. Их надо брать положительными. Но увеличение x_5 повлечет за собой увеличение значения линейной формы. Следовательно, в новом плане x_5 надо считать нулевым. В то же время x_1 нельзя сделать больше 2 (хотя его увеличение и вызывает уменьшение линейной формы), так как при $x_1 > 2$, и $x_5 = 0$ будет $x_3 < 0$, а при $x_1 = 2$ $x_3 = 0$. Положим $x_1 = 2$, $x_5 = 0$, тогда $x_3 = 0$, $x_2 = 5$, $x_4 = 4$, $x_6 = 12$. Получили все значения неизвестных неотрицательными, следовательно, опять имеем план

$$x = \begin{Vmatrix} 2 \\ 5 \\ 0 \\ 4 \\ 0 \\ 12 \end{Vmatrix}. \tag{48}$$

Значение линейной формы стало меньше: $F_1 = -39$.

Далее поступаем аналогично. Выбираем x_3 и x_5 свободными неизвестными и выражаем через них все остальные неизвестные и линейную форму:

$$\begin{cases} x_1 = 2 - \frac{1}{2} x_3 + \frac{1}{2} x_5, \\ x_2 = 5 - \frac{1}{3} x_5, \\ x_4 = 4 + x_3 - \frac{2}{3} x_5, \\ x_6 = 12 + \frac{3}{2} x_3 - \frac{3}{2} x_5; \end{cases} \tag{49}$$

$$F_1 = -39 + \frac{7}{2} x_3 - \frac{11}{6} x_5. \tag{50}$$

Здесь уменьшать значение линейной формы можно лишь путем увеличения значения x_5. Но при $x_3 = 0$ величину x_5 нельзя брать

больше чем 6, так как тогда будет $x_4 < 0$. При $x_5 = 6$ и $x_3 = 0$ имеем: $x_4 = 0$, $x_1 = 5$, $x_2 = 3$, $x_6 = 3$. Вновь получили все значения неизвестных неотрицательными. Поэтому имеем план

$$x = \begin{Vmatrix} 5 \\ 3 \\ 0 \\ 0 \\ 6 \\ 3 \end{Vmatrix}, \tag{51}$$

при котором значение линейной формы $F_1 = -50$ стало еще меньше, чем прежде.

Нулевые значения в плане (51) имеют неизвестные x_3 и x_4. Примем их за свободные неизвестные. Тогда получим

$$\begin{cases} x_1 = 5 + \frac{1}{4} x_3 - \frac{3}{4} x_4, \\ x_2 = 3 - \frac{1}{2} x_3 + \frac{1}{2} x_4, \\ x_5 = 6 + \frac{3}{2} x_3 - \frac{3}{2} x_4, \\ x_6 = 3 - \frac{3}{4} x_3 + \frac{9}{4} x_4 \end{cases} \tag{52}$$

$$F_1 = -50 + \frac{3}{4} x_3 + \frac{11}{4} x_4. \tag{53}$$

Теперь уже видим, что дальнейшее уменьшение линейной формы F_1 может быть достигнуто только за счет отрицательных значений x_3 и x_4, что, как известно, недопустимо. Следовательно, $F_1 = -50$ (при $x_3 = 0$ и $x_4 = 0$) есть в области существования решений наименьшее значение формы F_1. Так как первоначально нас интересовали неизвестные x_1 и x_2, то именно их значения мы выбираем из последнего (оптимального) плана: $x_1 = 5$; $x_2 = 3$. Задача решена. Ее результат, конечно, совпал с тем, что было получено в предыдущем параграфе геометрическим путем. Но здесь для решения задачи мы применили метод последовательного улучшения планов (с и м п л е к с - м е т о д).

Внимательно изучив все проделанное выше и обратившись опять к геометрическому представлению, замечаем, что составленные нами планы были связаны с крайними точками (вершинами) многоугольника решений, т. е. были опорными планами. Действительно, выбирая в качестве свободных неизвестных какие-либо две, мы именно их превращали в координатные переменные, а следовательно, их нулевые значения соответствовали в новой координатной системе началу координат, т. е. крайней точке многоугольника. Более того, нетрудно заметить (см. рис. 99), что при $x_1 = x_2 = 0$ опорный

план (44) соответствовал точке 0; при $x_1 = x_5 = 0$ опорный план (45) соответствовал точке M_1; при $x_3 = x_5 = 0$ опорный план (48) строился в точке M_2 и, наконец, при $x_3 = x_4 = 0$ опорный план (51) в точке M_3 оказался оптимальным.

Таким образом, улучшая опорные планы, мы передвигались последовательно от одной крайней точки (вершины) многоугольника решений к другой крайней точке, смежной с предыдущей, а опорный план оказался оптимальным лишь тогда, когда выбранные в последнем шаге свободные неизвестные вошли в состав линейной формы F_1 с неотрицательными коэффициентами. (Вполне понятно, что количество свободных неизвестных, как уже отмечалось выше, определяет размерность пространства, в котором проводится решение).

Именно эти два факта положены в основу симплексного метода, схема применения которого и записи вычислений в общем случае таковы:

1. Определяют общее количество n неизвестных, ранг r системы равенств-ограничений задачи и количество $k = n - r$ свободных неизвестных.

2. Выбирают k свободных неизвестных (они нумеруются первыми) $x_1, x_2, \ldots, x_k$ и через них выражают все остальные (базисные) r неизвестные $x_{k+1}, x_{k+2}, \ldots, x_{k+r} = x_n$ и линейную форму F. Получают систему равенств-ограничений и линейную форму задачи в виде

$$\begin{cases} x_{k+1} = \alpha'_{11}x_1 + \alpha'_{12}x_2 + \cdots + \alpha'_{1k}x_k + \beta_1, \\ x_{k+2} = \alpha'_{21}x_1 + \alpha'_{22}x_2 + \cdots + \alpha'_{2k}x_k + \beta_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n = \alpha'_{r1}x_1 + \alpha'_{r2}x_2 + \cdots + \alpha'_{rk}x_k + \beta_r \end{cases} \tag{54}$$

и

$$F = \gamma_0 + \gamma'_1x_1 + \gamma'_2x_2 + \cdots + \gamma'_kx_k. \tag{55}$$

Затем приводят запись системы (54) и линейной формы (55) к стандартному в симплекс-методе виду, где правые части уравнений представляют как разности между свободными членами и суммой остальных:

$$\begin{cases} x_{k+1} = \beta_1 - (-\alpha'_{11}x_1 - \alpha'_{12}x_2 - \cdots - \alpha'_{1k}x_k), \\ x_{k+2} = \beta_2 - (-\alpha'_{21}x_1 - \alpha'_{22}x_2 - \cdots - \alpha'_{2k}x_k), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n = \beta_r - (-\alpha'_{r1}x_1 - \alpha'_{r2}x_2 - \cdots - \alpha'_{rk}x_k) \end{cases} \tag{56}$$

и

$$F = \gamma_0 - (-\gamma'_1x_1 - \gamma'_2x_2 - \cdots - \gamma'_kx_k). \tag{57}$$

После введения в уравнения (56) и (57) обозначений

$-\alpha'_{ij} = \alpha_{ij}$ $(i = 1, 2, \ldots, k)$, $(j = 1, 2, \ldots, r)$ и $-\gamma'_i = \gamma_i$

стандартная форма записи системы равенств-ограничений и целевой функции F принимает следующий простой вид:

$$\begin{cases} x_{k+1} = \beta_1 - (\alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1k}x_k), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_{k+l} = \beta_l - (\alpha_{l1}x_1 + \alpha_{l2}x_2 + \cdots + \alpha_{lk}x_k), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_{k+l} = \beta_l - (\alpha_{l1}x_1 + \alpha_{l2}x_2 + \cdots + \alpha_{lk}x_k), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n = \beta_r - (\alpha_{r1}x_1 + \alpha_{r2}x_2 + \cdots + \alpha_{rk}x_k); \end{cases} \quad (58)$$

$$F = \gamma_0 - (\gamma_1 x_1 + \gamma_2 x_2 + \cdots + \gamma_k x_k). \quad (59)$$

3. Для упрощения дальнейших операций заполняют стандартную симплекс-таблицу, в которую заносят свободные члены и коэффициенты при неизвестных в уравнениях (58) и (59). При этом клетки таблицы делают достаточно большими и указанные числа записывают в верхних частях соответствующих клеток (каждую клетку удобно разделить на две части диагональю, идущей вверх слева направо или горизонтальной чертой), причем каждая строка соответствует уравнению системы ограничений, первая строка — форме F, первый столбец — свободным членам, остальные столбцы — коэффициентам при неизвестных (табл. 11).

Таблица 11

Функция F и базисные неизвестные	Свободные члены	Свободные неизвестные						
		x_1	...	x_S	...	x_j	...	x_k
F	γ_0	γ_1	...	γ_S	...	γ_j	...	γ_k
x_{k+1}	β_1	α_{11}	...	α_{1S}	...	α_{1j}	...	α_{1k}
...	...	...	...	...	...	...	...	...
x_{k+l}	β_l	α_{l1}	...	α_{lS}	...	α_{lj}	...	α_{lk}
...	...	...	...	...	...	...	...	...
x_{k+l}	β_l	α_{l1}	...	α_{lS}	...	α_{lj}	...	α_{lk}
...	...	...	...	...	...	...	...	...
x_n	β_r	α_{r1}	...	α_{rS}	...	α_{rj}	...	α_{rk}

4. С помощью симплекс-таблицы 11 выясняют, имеет ли вообще данная задача решение, и если да, то является ли выбранный план опорным.

Если все коэффициенты $\gamma_1, \gamma_2, \ldots, \gamma_k$, стоящие в первой строке таблицы, положительны, то это означает, что линейная форма F не ограничена снизу и оптимального решения задачи не существует. План решения будет опорным в случае, когда все свободные члены $\beta_1, \beta_2, \ldots, \beta_r$, стоящие в первом столбце таблицы, положительны.

5. Рассматривают те из свободных неизвестных, которые в скобках формы (59) стоят с положительными коэффициентами γ_i и среди них выбирают какую-либо одну (например, x_j). Если всем свободным неизвестным, кроме x_j, придать нулевые значения, а x_j увеличивать, то форма F будет уменьшаться. Свободную неизвестную x_j можно увеличивать до тех пор, пока впервые не станет нулем какая-нибудь базисная неизвестная.

Получается следующая картина.

Если все свободные неизвестные, кроме x_j, равны нулю, из системы (58) остается:

$$\begin{cases} x_{k+1} = \beta_1 - \alpha_{1j}x_j, \\ x_{k+2} = \beta_2 - \alpha_{2j}x_j, \\ \cdots\cdots\cdots \\ x_{k+l} = \beta_l - \alpha_{lj}x_j, \\ \cdots\cdots\cdots \\ x_n = \beta_r - \alpha_{rj}x_j. \end{cases} \tag{60}$$

Все базисные неизвестные $x_{k+\nu}$, у которых коэффициент $\alpha_{\nu j} = 0$, при увеличении x_j убывать не будут. Будут убывать, а, следовательно, могут обратится в нуль лишь те неизвестные, у которых коэффициент при x_j положителен.

Базисная неизвестная x_{k+l} обратится в нуль только при

$$x_j = \frac{\beta_l}{\alpha_{lj}} \quad (\alpha_{lj} > 0). \tag{61}$$

(Все отношения вида $\frac{\beta_l}{\alpha_{lj}}$ при условии $\alpha_{lj} > 0$ неотрицательны, так как все $\beta_l \geqslant 0$ по условию выбора свободных неизвестных).

Очевидно, среди всех базисных неизвестных, убывающих при возрастании от нуля свободной неизвестной x_j, в нуль обратится раньше всех та, для которой отношение вида (61) имеет наименьшее значение.

Пусть

$$\frac{\beta_i}{\alpha_{ij}} = \min\left(\frac{\beta_l}{\alpha_{lj}}\right) \quad (\alpha_{lj} > 0). \tag{62}$$

Тогда неизвестная x_{k+i} обратится в нуль раньше других, а остальные неизвестные все еще будут положительными.

Коэффициент α_{ij} важен для дальнейших расчетов, поэтому его называют генеральным (или «разрешающим» элементом) и подчеркивают в таблице. Обычно выделяют также жирными линиями строку и столбец, в которых стоит разрешающий элемент, и называют их разрешающей строкой и разрешающим столбцом.

Увеличивая значение x_j до величины $\frac{\beta_i}{\alpha_{ij}}$, перемещаются от исходной опорной точки (где все выбранные вначале свободные неизвестные, включая и x_j, были равны нулю) в новую опорную точку (где $x_j > 0$, а все остальные первоначальные свободные неизвестные и базисная неизвестная x_{k+i} равны нулю). При этом перемещении линейная форма F уменьшается от значения γ_0, до величины $\gamma_0 - \gamma_j \frac{\beta_i}{\alpha_{ij}}$.

С целью дальнейшего уменьшения значения линейной формы составляют новую симплекс-таблицу, в которой в качестве свободных неизвестных будут все прежние, а вместо x_j войдет x_{k+i}. Затем продолжают улучшение плана решения задачи, увеличивая значение одной из свободных неизвестных, для которой коэффициент γ_i в первой строке новой таблицы положительный.

6. Для подготовки новой симплекс-таблицы используют нижние части клеток предыдущей таблицы, куда записывают либо новые элементы (для клеток в разрешающих строке и столбце), либо соответствующие слагаемые новых элементов (для всех остальных клеток).

При составлении новой таблицы с помощью предшествующей руководствуются правилами, которые можно определить, исходя из следующего.

Чтобы осуществить вывод неизвестной x_j из числа свободных (сделать ее базисной), а вместо x_j в число свободных перевести x_{k+i} (бывшую базисную, первой обращающуюся в нуль при увеличении x_j), выражают все базисные неизвестные (включая и новую x_j), а также линейную форму F через свободные неизвестные (учитывая и новую свободную неизвестную x_{k+i}).

Так как x_{k+i} выходит из числа базисных, то именно из уравнения

$$x_{k+i} = \beta_i - (\alpha_{i1}x_1 + \alpha_{i2}x_2 + \cdots + \alpha_{ij}x_j + \cdots + \alpha_{ik}x_k)$$

находят x_j:

$$x_j = \frac{1}{\alpha_{ij}}(\beta_i - (\alpha_{i1}x_1 + \alpha_{i2}x_2 + \cdots + x_{k+i} + \cdots + \alpha_{ik}x_k)) =$$

$$= \beta_i\lambda - (\alpha_{i1}\lambda x_1 + \alpha_{i2}\lambda x_2 + \cdots + \lambda x_{k+i} + \cdots + \alpha_{ik}\lambda x_k), \quad (63)$$

где $\lambda = \frac{1}{\alpha_{ij}}$.

Это значение x_j подставляют во все равенства, выражающие базисные неизвестные и форму F. После несложных преобразований

получают зависимости вида

$$x_{k+t} = (\beta_t - \lambda\beta_i\alpha_{tj}) - \left(\sum_{S=1}^{k} (\alpha_{tS} - \lambda\alpha_{tj}\alpha_{iS})\, x_S - \lambda\alpha_{tj}x_{k+t}\right), \quad (64)$$

где $t = 1, 2, \ldots, i-1, i+1, \ldots, r$ и

$$F = (\gamma_0 - \lambda\beta_i\gamma_j) - \left(\sum_{S=1}^{k} (\gamma_S - \lambda\alpha_{iS}\gamma_j)\, x_S - \lambda\gamma_j x_{k+t}\right). \quad (65)$$

Рассмотрев выражения (64) и (65), можно сформулировать следующие правила:

Для составления новой симплекс-таблицы при переходе от x_j к x_{k+i} надо:

разрешающий элемент заменить на обратную ему величину $\left(т.\ е.\ на\ \lambda = \frac{1}{\alpha_{ij}}\right)$;

Таблица 12

		x_1	...	x_S	...	x_j	...	x_k
F	γ_0 / $-\lambda\beta_i\gamma_j$	γ_1 / $-\lambda\alpha_{i1}\gamma_j$	...	γ_S / $-\lambda\alpha_{iS}\gamma_j$	...	γ_j / $-\lambda\gamma_j$	...	γ_k / $-\lambda\alpha_{ik}\gamma_j$
x_{k+1}	β_1 / $-\lambda\beta_i\alpha_{1j}$	α_{11} / $-\lambda\alpha_{i1}\alpha_{1j}$	...	α_{1S} / $-\lambda\alpha_{iS}\alpha_{1j}$	...	α_{1j} / $-\lambda\alpha_{1j}$	...	α_{1k} / $-\lambda\alpha_{ik}\alpha_{1j}$
...	...	...	...	...	...	...	...	...
x_{k+i}	β_i / $\lambda\beta_i$	α_{i1} / $\lambda\alpha_{i1}$	...	α_{iS} / $\lambda\alpha_{iS}$	...	α_{ij} / λ	...	α_{ik} / $\lambda\alpha_{ik}$
...	...	...	...	...	...	...	...	...
x_{k+l}	β_l / $-\lambda\beta_i\alpha_{lj}$	α_{l1} / $-\lambda\alpha_{i1}\alpha_{lj}$	...	α_{lS} / $-\lambda\alpha_{iS}\alpha_{lj}$	...	α_{lj} / $-\lambda\alpha_{lj}$	...	α_{lk} / $-\lambda\alpha_{ik}\alpha_{lj}$
...	...	...	...	...	...	...	...	...
x_n	β_r / $-\lambda\beta_i\alpha_{rj}$	α_{r1} / $-\lambda\alpha_{i1}\alpha_{rj}$	...	α_{rS} / $-\lambda\alpha_{iS}\alpha_{rj}$	...	α_{rj} / $-\lambda\alpha_{rj}$	...	α_{rk} / $-\lambda\alpha_{ik}\alpha_{rj}$

Таблица 13

		x_1	...	x_S	...	x_{k+i}	...	x_k
F	$\gamma_0-\lambda\beta_i\gamma_i$	$\gamma_1-\lambda\alpha_{i1}\gamma_j$	...	$\gamma_S-\lambda\alpha_{iS}\gamma_j$	...	$-\lambda\gamma_j$	...	$\gamma_k-\lambda\alpha_{ik}\gamma_j$
x_{k+1}	$\beta_1-\lambda\beta_i\alpha_{1j}$	$\alpha_{11}-\lambda\alpha_{i1}\alpha_{1j}$	...	$\alpha_{1S}-\lambda\alpha_{iS}\alpha_{1j}$	...	$-\lambda\alpha_{1j}$	...	$\alpha_{1k}-\lambda\alpha_{ik}\alpha_{1j}$
...	...	...	...	...	...	...	...	...
x_j	$\lambda\beta_i$	$\lambda\alpha_{i1}$	...	$\lambda\alpha_{iS}$	...	λ	...	$\lambda\alpha_{ik}$
...	...	...	...	...	...	...	...	...
x_{k+l}	$\beta_l-\lambda\beta_i\alpha_{lj}$	$\alpha_{l1}-\lambda\alpha_{i1}\alpha_{lj}$	...	$\alpha_{lS}-\lambda\alpha_{iS}\alpha_{lj}$	...	$-\lambda\alpha_{lj}$	...	$\alpha_{lk}-\lambda\alpha_{ik}\alpha_{lj}$
...	...	...	...	..	...	...	...	...
x_n	$\beta_r-\lambda\beta_i\alpha_{rj}$	$\alpha_{r1}-\lambda\alpha_{i1}\alpha_{rj}$	...	$\alpha_{rS}-\lambda\alpha_{iS}\alpha_{rj}$	...	$-\lambda\alpha_{rj}$	...	$\alpha_{rk}-\lambda\alpha_{ik}\alpha_{rj}$

все остальные элементы разрешающей строки разделить на разрешающий элемент (умножить на λ);

все элементы разрешающего столбца, кроме самого разрешающего элемента, разделить на разрешающий элемент (умножить на λ) и перед каждым из них сменить знак на противоположный;

для каждой из клеток, не стоящей в разрешающих строке и столбце, получить произведение числа, стоящего в верхней части пересечения разрешающей строки со столбцом данной клетки, на число, стоящее в нижней части пересечения разрешающего столбца со строкой взятой клетки.

В табл. 12 с помощью этих правил заполнены нижние части клеток разрешающих строки и столбца новыми элементами, а всех остальных клеток — соответствующими слагаемыми новых элементов.

После применения перечисленных правил и использования слагаемых новых элементов клеток, не входящих в разрешающие строку и столбец, получается симплекс-таблица 13 для нового набора свободных неизвестных.

Теперь можно продолжать работу по той же схеме дальше, пока в результате конечного числа шагов не появится симплекс-таблица, в первой строке которой все коэффициенты γ_i, кроме γ_0, не станут отрицательными. Последнее будет означать, что получено оптимальное решение задачи линейного программирования, при котором линейная форма F минимизирована.

§ 47. ПОНЯТИЕ О ДИНАМИЧЕСКОМ ПРОГРАММИРОВАНИИ

Динамическое программирование является специальным вычислительным методом оптимизации решений задач, позволяющих представить процесс (операцию) в виде последовательности отдельных этапов (шагов).

В основе этого метода лежит принцип оптимальности, утверждающий, что каков бы ни был путь достижения некоторого состояния системы, последующие решения должны принадлежать оптимальной программе для оставшейся части пути, начинающейся с этого состояния. Применение принципа оптимальности дает возможность получить все используемые в динамическом программировании функциональные (рекуррентные) соотношения [1].

Метод динамического программирования пригоден не только для дискретных систем, но и для получения приближенных решений непрерывных задач оптимизации путем приведения последних к дискретной форме. Как и метод линейного программирования, метод динамического программирования хорошо «приспособлен» к использованию современных средств вычислительной техники.

В данном параграфе рассматриваются лишь основные положения динамического программирования, которые иллюстрируются простейшим примером, поясняющим этот важнейший метод решения задач оптимального планирования (управления).

Общая постановка задачи динамического программирования. Рассмотрим некоторую физическую систему S, которая с течением времени может изменять свое состояние. Допустим, что мы можем влиять на состояние системы, переводя ее из одного состояния в другое. Назовем такую систему управляемой, а действия, с помощью которых мы влияем на состояние системы,— управлением. Обозначим управление буквой u.

Предположим, что при изменении состояния системы S изменяется некоторая величина W, которую будем называть критерием.

Критерий W зависит от управления u

$$W = W(u). \tag{66}$$

Пусть $\tilde{S}_0$ — область начальных состояний, $\tilde{S}_{\text{кон}}$ — область конечных состояний системы S.

Поставим следующую задачу оптимального управления:

Из множества возможных управлений u найти такое управление u^*, которое переводит систему S из начального состояния $S_0 \in \tilde{S}_0$ в конечное состояние $S_{\text{кон}} \in \tilde{S}_{\text{кон}}$ так, чтобы критерий $W(u)$ обращался в максимум (иногда представляет интерес случай, когда критерий W обращается в минимум).

---

[1] См.: Р. Беллман, И. Г. Гликсберг, О. Гросс. Некоторые вопросы математической теории процессов управления. М., изд-во Иностр. лит., 1962, с. 183.

Если W^* — наибольшее значение критерия W при всех возможных управлениях u, то записывают

$$W^* = \max_u (W(u)). \qquad (67)$$

Запись $\max_u$ читается «максимум по u» и означает: «максимальное из всех значений $W(u)$ при всех возможных управлениях u». То из управлений, при котором достигается этот максимум, и есть оптимальное управление u^*[1].

Состояние физической системы S всегда можно охарактеризовать с помощью того или иного количества параметров. Эти параметры называются фазовыми координатами системы. Состояние системы изображается точкой с этими координатами в так называемом фазовом пространстве. Изменение состояния системы в процессе управления изображается как перемещение точки в этом фазовом пространстве, а выбор управления u представляет выбор определенной траектории точки в фазовом пространстве — определенного закона движения этой точки.

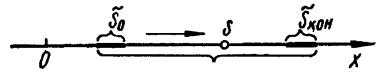

Рис. 101

Если состояние системы S характеризуется одним параметром — координатой x, то изменение этой координаты изображается перемещением точки по оси Ox или по некоторому ее участку. В этом случае фазовое пространство одномерное, а управление u представляет собой закон движения точки S из исходного состояния $S_0 \in \tilde{S}_0$ в конечное состояние $S_{кон} \in \tilde{S}_{кон}$ (рис. 101).

Если состояние системы S характеризуется двумя параметрами x_1 и x_2, то фазовым пространством будет плоскость x_1Ox_2 или некоторая ее часть, а управление u будет изображаться перемещением точки S из области $\tilde{S}_0$ в область $\tilde{S}_{кон}$ (рис. 102).

В общем случае, когда состояние системы S определяется n параметрами $x_1, x_2, \ldots, x_n$, его изображают точкой в n-мерном пространстве.

Теперь задачу оптимального управления можно сформулировать так:

Найти такое управление u^* (оптимальное управление), при котором точка S фазового пространства переместится из начальной области $\tilde{S}_0$ в конечную область $\tilde{S}_{кон}$ так, что при этом критерий W обратится в максимум (минимум).

Существуют различные методы решения задачи оптимального управления.

Рассмотрим ее решение методом динамического программирования.

---

[1] См.: Е. С. Вентцель. Исследование операций. М., Советское радио, 1972, с. 133.

Процесс перемещения точки S из $\tilde{S}_0$ в $\tilde{S}_{кон}$ разобъем на несколько этапов (шагов), и на каждом из них, начиная с последнего, найдем условное оптимальное управление (при всевозможных предположениях о результатах предыдущего шага).

После того как процесс оптимизации доведен до исходного состояния S_0, снова проходится вся последовательность шагов, начиная с первого, и на каждом шаге из множества условных оптимальных управлений выбирается одно (понятно, что на первом шаге оптимальное управление будет безусловным).

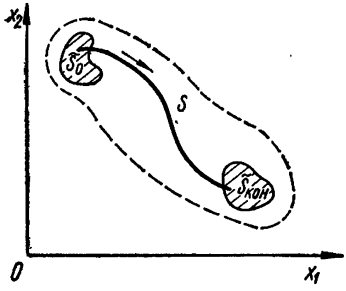

Рис. 102

Предположим, что критерий (66) является аддитивным:

$$W = \sum_{i=1}^{m} W_i, \tag{68}$$

где W_i — приращение критерия W на i-м шаге.

Ознакомимся с общей схемой решения задачи динамического программирования с аддитивным критерием (68).

Предположим, что процесс управления физической системы S расчленен на m шагов. На каждом i-м шаге имеется управление u_i, с помощью которого система из состояния S_{i-1}, в котором она оказалась в результате $(i-1)$-го шага, переходит в состояние S_i.

Для применения метода динамического программирования существенно, чтобы новое состояние S_i зависело только от состояния S_{i-1} и управления u_i на i-м шаге и не зависело от того, каким образом система пришла в состояние S_{i-1}:

$$S_i = S_i(S_{i-1}, u_i). \tag{69}$$

Под действием управлений $u_1, u_2, \ldots, u_m$ система переходит из начального состояния S_0 в конечное состояние $S_{кон}$. В результате за m шагов критерий W примет значение

$$W = \sum_{i=1}^{m} W_i(S_{i-1}, u_i), \tag{70}$$

где $W_i(S_{i-1}, u_i)$ — приращение критерия W на i-м шаге, зависящее как от предыдущего состояния системы S_{i-1}, так и от управления u_i на i-м шаге.

Надо выбрать начальное состояние системы $S_0 \in \tilde{S}_0$ и управление $u_1^*, u_2^*, \ldots, u_m^*$ на каждом шаге так, чтобы после m шагов система перешла в конечное состояние $S_{кон} \in \tilde{S}_{кон}$, а критерий W обратился при этом в максимум (минимум), т. е. чтобы выполнялось условие (67).

Введем обозначения:

W_m — приращение критерия W за последний шаг,

$W_{m-1,m}$ — приращение критерия W за два последних шага,

. . . . . .

$W_{i,i+1,\ldots,m}$ — приращение критерия W за последние $(m-i+1)$ шага, начиная с i-го и кончая шагом с номером m.

При этом

$$W_{i,i+1,\ldots,m} = W_i + W_{i+1} + \cdots + W_m.$$

Пусть после $(m-1)$-го шага система находится в состоянии S_{m-1}. Так как последний шаг должен привести систему в состояние $S_m = S_{\text{кон}} \in \tilde{S}_{\text{кон}}$, то в качестве состояний S_{m-1} можно брать лишь такие, из которых за один шаг можно перейти в область $\tilde{S}_{\text{кон}}$.

Допустим, что состояние S_{m-1} нам известно. Найдем при этом условии условное оптимальное управление $u_m^*(S_{m-1})$, которое переводит систему в конечное состояние $S_m = S_{\text{кон}} \in \tilde{S}_{\text{кон}}$ так, что приращение W_m критерия W на этом последнем шаге имеет максимальное значение

$$W_m^*(S_{m-1}) = \max_{u_m} (W_m(S_{m-1}, u_m)).$$

Перейдем теперь к оптимизации предпоследнего $(m-1)$-го шага. Предположим, что в результате $(m-2)$-го шага система пришла в состояние S_{m-2}. В результате управления u_{m-1} система из состояния S_{m-2} перейдет в состояние S_{m-1}, а критерий W получит при этом приращение $W_{m-1} = W_{m-1}(S_{m-2}, u_{m-1})$, зависящее как от состояния системы S_{m-2}, так и от управления u_{m-1} на $(i-1)$-м шаге. Состояние системы S_{m-1} тоже зависит от S_{m-2} и управления u_{m-1}:

$$S_{m-1} = S_{m-1}(S_{m-2}, u_{m-1}).$$

Но для любого состояния S_{m-1} следующий шаг с номером m уже оптимизирован, и максимальное приращение критерия W на m-м шаге

$$W_m^*(S_{m-1}) = W_m^*(S_{m-1}(S_{m-2}, u_{m-1})),$$

при этом $W_m^*(S_{m-1}) \sim u_m^*(S_{m-1})$ (это означает, что величине $W_m^*(S_{m-1})$ соответствует условное оптимальное управление $u_m^*(S_{m-1})$).

Пусть $W_{m-1,m}^+$ — приращение критерия W на двух последних шагах с номерами $m-1$ и m при любом управлении на $(m-1)$-м шаге и оптимальном управлении на m-м шаге:

$$W_{m-1,m}^+ = W_{m-1}(S_{m-2}, u_{m-1}) + W_m^*(S_{m-1}(S_{m-2}, u_{m-1})).$$

Выбираем такое оптимальное условное управление $u_{m-1}^*(S_{m-2})$ на $(m-1)$-м шаге, при котором величина $W_{m-1,m}^+$ достигает максимума:

$$W_{m-1,m}^*(S_{m-2}) = \max_{u_{m-1}} (W_{m-1,m}^+(S_{m-2}, u_{m-1})).$$

При этом

$$W^*_{m-1,m}(S_{m-2}) \sim u^*_{m-1}(S_{m-2}).$$

Конечно, предполагаем, что в качестве состояний системы S_{m-2} после $(m-2)$-х шагов берутся лишь те, из которых можно за два шага перейти в $\tilde{S}_{\text{кон}}$.

Аналогично можно найти условное оптимальное управление на $(m-2)$-м шаге и максимальное значение

$$W^*_{m-2,m-1,m}(S_{m-3}) \sim u^*_{m-2}(S_{m-3}),$$

где

$$W^*_{m-2,m-1,m}(S_{m-3}) = \max_{u_{m-2}} (W^+_{m-2,m-1,m}(S_{m-3}, u_{m-2})),$$

$$W^+_{m-2,m-1,m} = W_{m-2}(S_{m-3}, u_{m-2}) + W^*_{m-1,m}(S_{m-2}(S_{m-3}, u_{m-2}))$$

есть сумма приращений критерия W при любом управлении на $(m-2)$-м шаге и оптимальных управлениях на $(m-1)$-м и m-м шагах.

Применяя последовательно такой процесс, дойдем до первого шага. На первом шаге надо выбрать оптимальным образом начальное состояние системы S^*_0. При этом могут быть два случая:

1. Начальное состояние S_0 в точности задано, вся область $\tilde{S}_0$ состоит из одной точки S_0. Тогда $S^*_0 = S_0$.

2. Начальное состояние можно брать любым на области $\tilde{S}_0$. В этом случае нужно оптимизировать выбор начального состояния S_0 так, чтобы получить максимум величины $W_{1,\ldots,m}(S_0)$:

$$W^*_{1,\ldots,m} = \max_{S^*_0 \in \tilde{S}_0} (W^*_{1,\ldots,m}(S_0)).$$

Остается найти оптимальное управление в окончательной форме (пока найдены лишь условные оптимальные управления на всех шагах, кроме первого; на первом шаге найдено безусловное оптимальное управление).

В качестве начального состояния системы берем S^*_0, указанное выше. На первом шаге выбираем оптимальное управление $u^*_1 = u^*_1(S^*_0)$. После этого система переходит в новое состояние $S^*_1 = S_1(S^*_0, u^*_1)$. Оптимальное управление $u^*_2(S_1)$ на втором шаге найдено, причем для любого результата первого шага. Подставив S^*_1 вместо S_1, получим

$$u^*_2 = u^*_2(S^*_1).$$

Продолжая этот процесс, найдем безусловное оптимальное управление на последнем шаге:

$$u^*_m = u^*_m(S^*_{m-1}).$$

Окончательно получим оптимальное управление

$$u^* = (u_1^*, u_2^*, \ldots, u_m^*),$$

состоящее из оптимальных управлений на всех отдельных шагах, и максимальное значение критерия W^*.

Принятая в динамическом программировании система символических формул, которая использована в данном параграфе, непригодна для непосредственных вычислений. Однако эта система облегчает понимание сущности метода и помогает организовать процесс самого программирования.

Для лучшего уяснения идеи метода динамического программирования рассмотрим пример.

Задача о наборе высоты и скорости[1]. Летательный аппарат находится на высоте H_0 и имеет скорость v_0. Надо перевести его на высоту $H_{\text{кон}} > H_0$, получив при этом скорость $v_{\text{кон}} > v_0$. Известен расход горючего для подъема аппарата с любой высоты на другую высоту при заданной скорости и расход горючего для увеличения скорости при постоянной высоте полета. Требуется найти оптимальный режим набора высоты и увеличения скорости, при котором расход горючего W будет наименьшим.

Для того чтобы облегчить понимание метода динамического программирования, будем решать задачу в наиболее простой форме: предположим, что процесс набора высоты и скорости состоит из нескольких шагов, на каждом из которых летательный аппарат увеличивает или только высоту или только скорость.

Состояние летательного аппарата определяется параметрами v, H. Фазовое пространство — двумерное (плоскость vOH).

Фазовая траектория, переводящая точку S, изображающую летательный аппарат на фазовой плоскости vOH, из положения S_0 (v_0, H_0) в положение $S_{\text{кон}}$ ($v_{\text{кон}}$, $H_{\text{кон}}$) в данном случае является некоторой ломаной.

Разобъем приращение высоты ($H_{\text{кон}} - H_0$) на n_1 равных частей:

$$\Delta H = \frac{H_{\text{кон}} - H_0}{n_1},$$

приращение скорости ($v_{\text{кон}} - v_0$) — на n_2 равных частей:

$$\Delta v = \frac{v_{\text{кон}} - v_0}{n_2}.$$

Тогда весь процесс набора высоты и скорости будет состоять из $m = n_1 + n_2$ шагов. Точка S из S_0 в $S_{\text{кон}}$ может перемещаться по горизонтальным и вертикальным отрезкам. Запишем на каждом из этих отрезков соответствующий расход горючего. Каждой траектории, переводящей точку S из S_0 в $S_{\text{кон}}$, соответствует свой расход горючего W. Надо из всех траекторий выбрать оптимальную, на которой расход горючего будет наименьшим.

---

[1] См.: Е. С. Вентцель. Элементы динамического программирования. М., Наука, 1964.

Возьмем для простоты $n_1 = n_2 = 4$, тогда $m = 8$ (рис. 103). Займемся оптимизацией последнего (восьмого) шага. Рассмотрим правый верхний угол прямоугольной сетки с конечной точкой $S_{кон}$ (рис. 104).

В конце седьмого шага (шага с номером $m - 1$) точка S, изображающая летательный аппарат на фазовой плоскости, может оказаться либо в точке S_7^1, либо в точке S_7^2, из которых за один последний шаг можно перейти в точку $S_{кон}$, причем единственным способом. Поэтому выбора условного управления на восьмом шаге нет — оно для каждой точки S_7^1 и S_7^2 единственно. Около каждой из этих точек в кружке напишем минимальные (в данном случае неизбежные) расходы горючего, которые необходимо затратить для того, чтобы из них попасть в точку $S_{кон}$. Стрелкой около кружков обозначим направление, по которому должна двигаться точка S, если она окажется в этой позиции.

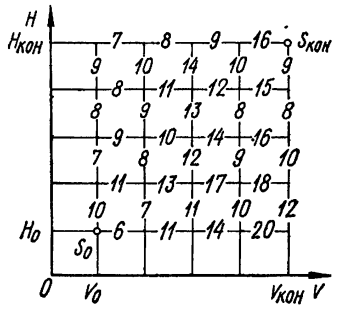

Рис. 103

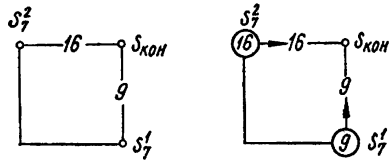

Рис. 104

Перейдем к выбору оптимального управления на предпоследнем (седьмом) шаге. Для этого надо рассмотреть все исходы предыдущего для него (шестого) шага, из которых за два шага можно попасть в точку $S_{кон}$. Таких исходов три — точки S_6^1, S_6^2, S_6^3 (рис. 105). Переход из точек S_6^1 и S_6^3 в точку $S_{кон}$ за два шага можно выполнить лишь единственным способом. В кружках около этих точек стоят цифры 17 и 25, показывающие расход горючего, необходимый для перехода из них в точку $S_{кон}$, стрелки около кружков указывают направления, по которым должна двигаться точка S, если она окажется в этих точках. Из точки S_6^2 в точку $S_{кон}$ за два шага можно попасть двумя способами: направо — вверх или вверх — направо. В первом случае потребуется $15 + 9 = 24$ единицы горючего, во втором случае $10 + 16 = 26$ единиц. Наименьший расход горючего будет при движении направо — вверх (его отмечаем стрелкой). Условное оптимальное управление для состояния S_6^2 на шестом шаге заключается в увеличении скорости на величину Δv при постоянной высоте $H_0 + 3\Delta H$.

Если бы расход горючего на том и другом перемещениях из S_6^2 в $S_{кон}$ оказался одним и тем же, за оптимальное управление на шестом шаге из состояния S_6^2 можно было бы принять любое из этих

двух возможных. Далее аналогичным образом оптимизируем пятый шаг, рассматривая все точки, из которых за три шага можно перейти в точку $S_{кон}$.

Продолжая этот процесс, дойдем до начальной точки S_0; получим схему (рис. 106), где в каждом кружке написан наименьший расход горючего при перемещении из этой точки в точку $S_{кон}$; стрелка показывает направление, в котором надо перемещаться из этой точки для того, чтобы расход горючего был наименьшим.

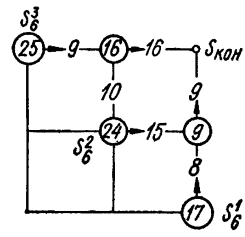

Рис. 105

Рассматривая схему, приведенную на рис. 106, заключаем, что оптимальным управлением будет следующее: на первом шаге — увеличение скорости на Δv при постоянной высоте H_0; на 2, 3, 4 и 5 шагах — набор высоты до $H_{кон}$ при постоянной скорости $v_0 + \Delta v$; на 6, 7 и 8 шагах — увеличение скорости до $v_{кон}$ при постоянной высоте $H_{кон}$.

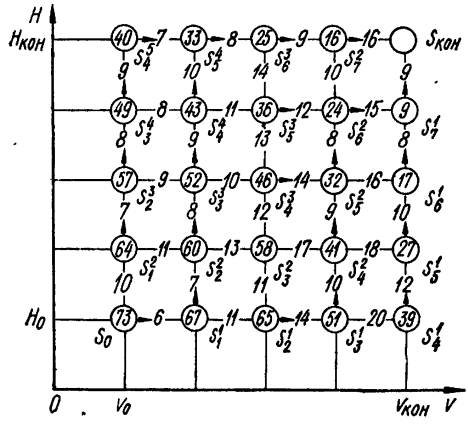

Рис. 106

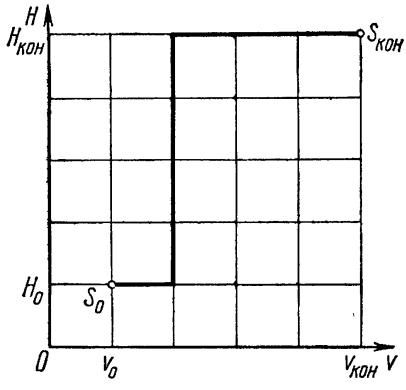

Рис. 107

При таком управлении будет наименьший расход горючего (в данном примере 73 условные единицы).

На рис. 107 показана фазовая траектория, соответствующая оптимальному управлению.

ГЛАВА IX

ПРИБЛИЖЕННЫЕ ВЫЧИСЛЕНИЯ

Во многих случаях применение математики к решению задач физики, техники, механики, экономики и других областей связано с выполнением приближенных вычислений. Это обусловлено не только тем, что исходные числовые данные о реальных объектах

редко бывают точными значениями соответствующих величин, или тем, что формулы, описывающие явления, оказываются приближенными, но также и тем, что для практических приложений часто нет необходимости в получении точных численных результатов. Вместе с тем, решение сложных задач, приводящих к огромному количеству операций, с помощью современных средств вычислительной техники требует использования приближенных алгоритмов, содержащих конечное число действий.

Инженер любой специальности всегда должен получить количественный результат и поэтому не может обойтись без знаний основ приближенных вычислений. Это в полной мере относится и к инженеру электро- и радиотехнической специальности.

В данной главе изложены лишь основные сведения по приближенным вычислениям, которые крайне необходимы как при использовании простейших счетно-решающих устройств, так и современных ЭВМ. В конце главы рассмотрены некоторые приближенные методы расчета переходных процессов в нелинейных электрических цепях.

§ 48. ПРИБЛИЖЕННЫЕ ЧИСЛА. ПОГРЕШНОСТИ

Приближенным числом a называется число, которое мало отличается от точного A и заменяет его в вычислениях.

Разность $\Delta a = A - a$ называется погрешностью приближенного числа a.

Величина $\Delta = |\Delta a| = |A - a|$ называется абсолютной погрешностью приближенного числа a.

Так как часто точное число A неизвестно, то найти абсолютную погрешность Δ не всегда возможно. Поэтому рассматривают предельную абсолютную погрешность Δ_a, где Δ_a — любое положительное число, такое, что $\Delta \leqslant \Delta_a$.

Тогда

$$|A - a| \leqslant \Delta_a,$$

$$-\Delta_a \leqslant A - a \leqslant \Delta_a, \quad a - \Delta_a \leqslant A \leqslant a + \Delta_a.$$

Иногда записывают $A = a \pm \Delta_a$.

Число $\delta = \dfrac{\Delta}{|A|}$ называется относительной погрешностью приближенного числа a.

Абсолютная погрешность $\Delta = \delta |A|$.

Любое положительное число δ_a, такое, что $\delta \leqslant \delta_a$, называется предельной относительной погрешностью приближенного числа a. Очевидно, $\Delta_a = \delta_a |A|$.

Так как обычно A мало отличается от a, то полагают

$$\Delta_a = |a| \delta_a,$$

тогда условно записывают: $A = a(1 \pm \delta_a)$.

Наиболее часто приближенное число a записывают в виде десятичной дроби.

Значащей цифрой приближенного числа называется любая цифра в его десятичной записи, отличная от нуля, и нуль, стоящий между значащими цифрами или являющийся представителем сохраненного десятичного разряда.

Например, в числе 3804 все цифры, включая и нуль,— значащие. В числе 0,0175 значащими цифрами будут 1,7 и 5. В числе 28,30 все четыре цифры — значащие, так как здесь цифра 0 показывает, что в данном приближенном числе сохранен десятичный разряд 10^{-2}.

Если абсолютная погрешность приближенного числа не превышает половины единицы разряда, выраженного n-й значащей цифрой (считая слева направо), то n первых значащих цифр в записи этого числа считаются верными. В математических таблицах все помещенные значащие цифры являются верными.

Точность приближенного числа зависит не от количества значащих цифр, а от количества верных значащих цифр. Если приближенное число содержит излишнее количество неверных значащих цифр, прибегают к округлению.

Обычно при выполнении приближенных вычислений число значащих цифр промежуточных результатов берется больше числа верных знаков на одну или две единицы. Окончательный результат может содержать не более чем одну излишнюю цифру по сравнению с верными.

Можно показать, что если положительное приближенное число a имеет n верных десятичных знаков, то относительная погрешность этого числа

$$\delta \leqslant \frac{1}{\alpha_m}\left(\frac{1}{10}\right)^{n-1},$$

где α_m — первая значащая цифра числа a.

Пример. Сколько десятичных знаков надо взять при вычислении $\sqrt{22}$, чтобы относительная погрешность не превышала 0,1%?

Решение. Здесь первая значащая цифра 4, $\alpha_m = 4$, $\delta = 0{,}001$. Поэтому

$$\frac{1}{4 \cdot 10^{n-1}} \leqslant 0{,}001;\ 10^{n-1} \geqslant 250,\ n \geqslant 4.$$

Погрешность суммы и разности. *Абсолютная погрешность алгебраической суммы нескольких приближенных чисел не превышает суммы абсолютных погрешностей этих чисел.*

Пусть $x_1, x_2, \ldots, x_n$ — приближенные числа. Рассмотрим их алгебраическую сумму

$$u = x_1 \pm x_2 \pm \cdots \pm x_n.$$

Тогда

$$\Delta u = \Delta x_1 \pm \Delta x_2 \pm \cdots \pm \Delta x_n,$$

$$|\Delta u| \leqslant |\Delta x_1| + |\Delta x_2| + \cdots + |\Delta x_n|.$$

За предельную абсолютную погрешность алгебраической суммы можно принять сумму предельных абсолютных погрешностей слагаемых.

Предельная абсолютная погрешность суммы не может быть меньше наибольшей предельной абсолютной погрешности слагаемых. В связи с этим применяют следующее правило сложения приближенных чисел:

1) выделяют числа, десятичная запись которых обрывается раньше, чем у других;

2) остальные числа округляют, сохраняя один или два запасных десятичных знака;

3) складывают полученные приближенные числа, учитывая все сохраненные знаки;

4) полученный результат округляют на один знак.

Если слагаемые одного и того же знака, то предельная относительная погрешность их суммы не превышает наибольшей из предельных относительных погрешностей слагаемых.

Пусть $u = x_1 - x_2$, тогда $\Delta u = \Delta_{x_1} + \Delta_{x_2}$, $\delta_u = \dfrac{\Delta_{x_1} + \Delta_{x_2}}{A}$, где A — точное значение абсолютной величины разности чисел x_1 и x_2.

Если x_1 и x_2 достаточно близки друг к другу и имеют малые абсолютные погрешности, то число A мало.

В этом случае предельная относительная погрешность разности может быть весьма большой, в то время как относительные погрешности вычитаемого и уменьшаемого малы. Поэтому при приближенных вычислениях следует по возможности избегать вычитания двух почти равных приближенных чисел. Если этого избежать нельзя, то надо уменьшаемое и вычитаемое брать с достаточным числом запасных знаков.

Погрешность произведения. Относительная погрешность произведения нескольких приближенных чисел, отличных от нуля, не превышает суммы относительных погрешностей этих чисел.

Пусть $x_1, x_2, \ldots, x_n$ — положительные приближенные числа, $u = x_1 \cdot x_2 \cdot \ldots \cdot x_n$. Тогда

$$\ln u = \ln x_1 + \ln x_2 + \cdots + \ln x_n;$$

$$\frac{\Delta u}{u} = \frac{\Delta x_1}{x_1} + \frac{\Delta x_2}{x_2} + \cdots + \frac{\Delta x_n}{x_n};$$

$$\left|\frac{\Delta u}{u}\right| \leqslant \left|\frac{\Delta x_1}{x_1}\right| + \left|\frac{\Delta x_2}{x_2}\right| + \cdots + \left|\frac{\Delta x_n}{x_n}\right|,$$

но

$$\left|\frac{\Delta x_l}{x_l}\right| \approx \left|\frac{\Delta x_l}{A_l}\right| = \delta_l, \quad \left|\frac{\Delta u}{u}\right| = \delta,$$

где A_i — точное значение приближенного числа x_i; δ_i — относительная погрешность.

Следовательно,

$$\delta \leqslant \delta_1 + \delta_2 + \cdots + \delta_n.$$

Эта формула будет верной и в случае любых знаков сомножителей $x_1, x_2, \ldots, x_n$.

Можно убедиться в том, что если все сомножители имеют m верных десятичных знаков (а число их не больше 10), то число верных знаков произведения на одну или на две единицы меньше m.

Если сомножители обладают различной точностью, то под m следует понимать число верных знаков в наименее точном из сомножителей.

Предельную относительную погрешность произведения n сомножителей ($n \leqslant 10$) можно найти по формуле

$$\delta_u = \frac{1}{2}\left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2} + \cdots + \frac{1}{\alpha_n}\right) \cdot \left(\frac{1}{10}\right)^{m-1},$$

где $\alpha_1, \alpha_2, \ldots, \alpha_n$ — первые значащие цифры в десятичной записи множителей.

Пример. Найти предельную относительную погрешность и число верных цифр произведения

$$u = 15{,}36 \cdot 12{,}582.$$

Решение. В данном случае $\alpha_1 = 1$, $\alpha_2 = 1$, $m = 4$.

$$\delta_u = \frac{1}{2}\left(\frac{1}{1} + \frac{1}{1}\right) \cdot \left(\frac{1}{10}\right)^{4-1} = 1 \cdot 10^{-3}.$$

В произведении будет не менее трех верных цифр.

Погрешность частного. *Относительная погрешность частного не превышает суммы относительных погрешностей делимого и делителя. При этом, если* $u = \frac{x}{y}$, *то* $\delta_u = \delta_x + \delta_y$.

Если делимое x и делитель y имеют по меньшей мере m верных знаков, а α и β — их первые значащие цифры, то предельную относительную погрешность частного можно найти по формуле

$$\delta_u = \frac{1}{2}\left(\frac{1}{\alpha} + \frac{1}{\beta}\right) \cdot \left(\frac{1}{10}\right)^{m-1}.$$

Если $\alpha \geqslant 2$ и $\beta \geqslant 2$, то частное имеет не менее $m - 1$ верных знаков.

Если $\alpha = 1$ или $\beta = 1$, то частное заведомо имеет $m - 2$ верных знаков.

§ 49. ПРИБЛИЖЕННОЕ РЕШЕНИЕ АЛГЕБРАИЧЕСКИХ И ТРАНСЦЕНДЕНТНЫХ УРАВНЕНИЙ

Пусть дано уравнение

$$f(x) = 0, \tag{1}$$

где функция $f(x)$ непрерывна в некотором $[a; b]$.

Как правило, редко удается найти корни этого уравнения. Поэтому большое значение имеют способы приближенного нахождения корней уравнения и оценки их точности.

Предположим, что уравнение (1) имеет лишь изолированные корни, т. е. для каждого корня существует окрестность, не содержащая других корней этого уравнения. Приближенное нахождение действительных корней уравнения обычно состоит из таких этапов:

1) Отделение корней — установление достаточно малых промежутков $[\alpha; \beta]$, содержащих один корень.

2) Уточнение приближенных корней с учетом заданной степени точности.

Для отделения корней используется теорема:

Если в $[\alpha; \beta]$ *функция* $f(x)$ *непрерывна, ее производная* $f'(x)$ *в* $]\alpha; \beta[$ *сохраняет знак и* $f(\alpha) \cdot f(\beta) < 0$, *то в* $]\alpha; \beta[$ *имеется один корень уравнения* $f(x) = 0$.

Надо помнить, что алгебраическое уравнение степени n имеет не более n действительных корней. Если для такого уравнения получили $n + 1$ перемену знаков, то все его корни отделены.

Таблица 14

x	$f(x)$	x	$f(x)$
-2	$-$	$1{,}5$	$-$
0	$+$	3	$+$

Пример. Отделить корни уравнения

$$f(x) \equiv x^3 - 2x^2 - x + 2 = 0.$$

Р е ш е н и е. Составим таблицу (табл. 14).

Уравнение имеет три действительных корня в интервалах $]-2; 0[$, $]0; 1{,}5[$, $]1{,}5; 3[$.

Действительные корни уравнения $f(x) = 0$ приближенно можно определить графически как абсциссы точек пересечения графика функции $y = f(x)$ с осью Ox. Если уравнение заменить равносильным уравнением $\varphi(x) = \psi(x)$, то корни данного уравнения получим как абсциссы точек пересечения графиков функций $y = \varphi(x)$ и $y = \psi(x)$.

Для нахождения корня уравнения $f(x) = 0$ можно применить метод половинного деления.

Пусть $f(a) \cdot f(b) < 0$, $f'(x)$ сохраняет знак в $]a; b[$. Если $f\left(\frac{a+b}{2}\right) = 0$, то корень уравнения $\xi = \frac{a+b}{2}$. Если $f\left(\frac{a+b}{2}\right) \neq 0$, то выбираем ту из половин $\left[a; \frac{a+b}{2}\right]$, $\left[\frac{a+b}{2}; b\right]$ промежутка $[a; b]$, на концах которой $f(x)$ имеет разные знаки. Новый суженный промежуток

$[a_1; b_1]$ опять делим пополам и рассуждаем так же. В результате получим на некотором этапе или точный корень уравнения, или же последовательность вложенных друг в друга промежутков $[a_1; b_1], \ldots, [a_n; b_n]$ таких, что $f(a_n) \cdot f(b_n) < 0$ и $b_n - a_n = \frac{1}{2^n}(b - a)$.

За приближенное значение корня ξ можно взять a_n или b_n, при этом погрешность не будет превышать $\frac{1}{2^n}(b - a)$.

Метод хорд (способ пропорциональных частей). Пусть $f(a) < 0$, $f(b) > 0$. Вместо деления $[a; b]$ пополам разделим его в отношении $f(a) : f(b)$.

Тогда приближенное значение корня будет: $x_1 = a + h_1$, где

$$h_1 = -\frac{-f(a)}{-f(a)+f(b)}(b-a) = -\frac{f(a)}{f(b)-f(a)}(b-a).$$

Применив этот метод к тому из промежутков $[a; x_1]$, $[x_1; b]$, на концах которого $f(x)$ имеет противоположные знаки, найдем второе приближение корня x_2 и т. д.

Геометрически способ пропорциональных частей эквивалентен замене кривой $y = f(x)$ хордой, проходящей через точки $A(a, f(a))$ и $B(b, f(b))$ (рис. 108).

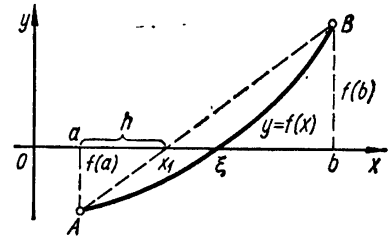

Рис. 108

Допустим, что $f''(x) > 0$ в $[a; b]$. Если при этом $f(a) > 0$, то конец a неподвижен, и последовательные приближения находятся по формулам

$$x_{n+1} = x_n - \frac{f(x_n)}{f(x_n)-f(a)}(x_n - a).$$

Если $f(a) < 0$, то неподвижен конец b, и последовательные приближения находятся по формулам

$$x_{n+1} = x_n - \frac{f(x_n)}{f(b)-f(x_n)}(b - x_n).$$

Случай, когда $f''(x) < 0$ в $[a; b]$, сводится к предыдущему, если рассматриваемое уравнение $f(x) = 0$ записать в виде: $-f(x) = 0$.

Вывод. *При применении метода хорд неподвижным будет тот конец промежутка $[a; b]$, где знак функции $f(x)$ совпадает со знаком ее второй производной $f''(x)$.*

Если найдены приближения x_{n+1} и x_n, то погрешность при замене корня ξ приближением x_n можно оценить по формуле

$$|\xi - x_n| \leqslant \frac{M_1 - m_1}{m_1}|x_n - x_{n-1}|,$$

где M_1 и m_1 — наибольшее и наименьшее значения модуля производной $f'(x)$ в $[a; b]$.

Если $M_1 \leqslant 2m_1$, то $|\xi - x_n| \leqslant |x_n - x_{n-1}|$.

Метод Ньютона (метод касательных). Пусть корень ξ уравнения $f(x) = 0$ находится в $[a; b]$, причем $f'(x)$ и $f''(x)$ сохраняют знаки в этом промежутке.

Предположим, что x_n — приближенное значение корня ξ.

Для его уточнения положим: $\xi = x_n + h_n$, где h_n — малая величина. Применив формулу Тейлора, получим

$$0 = f(\xi) = f(x_n + h_n) \approx f(x_n) + h_n f'(x_n).$$

Отсюда

$$h_n = -\frac{f(x_n)}{f'(x_n)}.$$

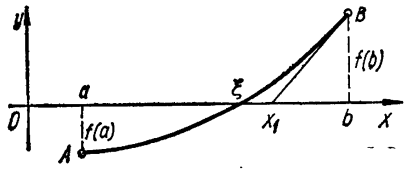

Рис. 109

Тогда следующее приближение корня ξ будет

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

Геометрически метод Ньютона означает замену дуги кривой $y = f(x)$ касательной, проведенной в одном из концов промежутка содержащего корень ξ. При этом касательную надо проводить в том конце промежутка, где знак функции совпадает со знаком ее второй производной (рис. 109).

Для оценки погрешности n-го приближения x_n можно применить формулу

$$|\xi - x_n| \leqslant \frac{|f(x_n)|}{m_1},$$

где m_1 — наименьшее значение $|f'(x)|$ в $[a; b]$.

Это можно сделать и по формуле

$$|\xi - x_n| \leqslant \frac{M_2}{2m_1} |x_n - b_{n-1}|^2,$$

где M_2 — наибольшее значение $|f''(x)|$ в $[a; b]$, m_1 — то же, что и выше.

Комбинированный метод. Пусть $f(a) \cdot f(b) < 0$, $f'(x)$ и $f''(x)$ сохраняют знаки в $[a; b]$.

Для нахождения приближенного значения корня ξ уравнения $f(x) = 0$ можно применить метод, представляющий комбинацию метода хорд и метода Ньютона.

Пусть, например, $f'(x) > 0$ и $f''(x) > 0$ в $[a; b]$.

Приближения, найденные по методу хорд, обозначим x_1, x_2, ..., x_n, ..., а приближения, найденные по методу Ньютона, обозначим $\bar{x}_1$, $\bar{x}_2$, ..., $\bar{x}_n$, ... Тогда

$$x_{n+1} = x_n - \frac{f(x_n)}{f(\bar{x}_n) - f(x_n)}(\bar{x}_n - x_n),$$

$$\bar{x}_{n+1} = \bar{x}_n - \frac{f(\bar{x}_n)}{f'(\bar{x}_n)},$$

$$x_{n+1} < \xi < \bar{x}_{n+1},$$

$$0 < \xi - x_{n+1} < \bar{x}_{n+1} - x_{n+1}.$$

Если допустимая абсолютная погрешность корня задана и равна ε, то вычисления следует закончить, когда $\bar{x}_{n+1} - x_{n+1} < \varepsilon$. При этом за приближенное значение корня обычно берут

$$\bar{\xi} = \frac{1}{2}(x_{n+1} + \bar{x}_{n+1}).$$

Остальные три случая комбинаций знаков у производных $f'(x)$ и $f''(x)$ в $[a; b]$ рассматриваются аналогично.

Метод итерации. Пусть дано уравнение

$$f(x) = 0, \tag{1}$$

где $f(x)$ — непрерывная функция. Требуется найти его вещественные корни.

Заменим это уравнение равносильным уравнением

$$x = \varphi(x). \tag{2}$$

(Это можно сделать различными способами, например, записав исходное уравнение в виде $x = f(x) + x$. В этом случае

$$\varphi(x) = f(x) + x).$$

Выберем каким-либо способом приближенное значение корня x_0 и подставим его в правую часть уравнения (2). Получим

$$x_1 = \varphi(x_0). \tag{3}$$

Повторяя этот процесс, находим

$$x_2 = \varphi(x_1), \ \ldots, \ x_n = \varphi(x_{n-1}). \tag{4}$$

Если существует $\lim\limits_{n\to\infty} x_n = \xi$, то, переходя в равенстве (4) к пределу и считая функцию $\varphi(x)$ непрерывной, получим

$$\lim_{n\to\infty} x_n = \varphi(\lim_{n\to\infty} x_{n-1}),$$

или

$$\xi = \varphi(\xi).$$

Следовательно, ξ — корень уравнения (2) (значит и корень уравнения (1)). Он может быть вычислен по формуле (4) с любой степенью точности.

Геометрически метод итерации изображен на рис. 110 и 111. Процесс итерации может оказаться и расходящимся (рис. 112). Поэтому полезна теорема, устанавливающая достаточные условия сходимости процесса итерации.

Пусть функция $\varphi(x)$ определена и дифференцируема в $[a; b]$, причем все ее значения $\varphi(x) \in [a; b]$. Тогда, если существует пра-

вильная дробь q такая, что $|\varphi'(x)| \leqslant q < 1$ при $a < x < b$, то:

1) процесс итерации $x_n = \varphi(x_{n-1})$ сходится независимо от начального значения $x_0 \in [a; b]$;

2) предельное значение $\xi = \lim\limits_{n\to\infty} x_n$ является единственным корнем уравнения $x = \varphi(x)$ в $[a; b]$.

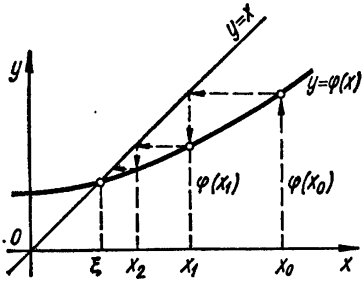

Рис. 110

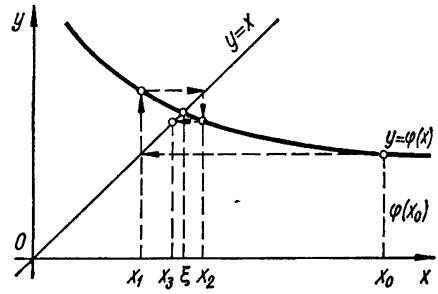

Рис. 111

Теорема остается правильной, если функция $\varphi(x)$ определена и дифференцируема в $]-\infty; +\infty[$. Для оценки приближения используем неравенства:

$$|\xi - x_n| \leqslant \frac{q^n}{1-q}|x_1 - x_0|,$$

$$|\xi - x_n| \leqslant \frac{|x_{n+1} - x_n|}{1-q},$$

$$|\xi - x_n| \leqslant \frac{q}{1-q}|x_n - x_{n-1}|.$$

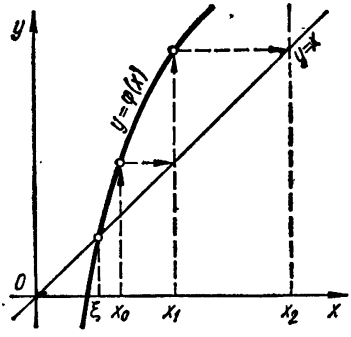

Рис. 112

Если $q < \frac{1}{2}$, то в этом случае из неравенства $|x_n - x_{n-1}| < \varepsilon$ вытекает неравенство $|\xi - x_n| < \varepsilon$.

Процесс итерации следует продолжать до тех пор, пока для двух последовательных приближений x_{n-1} и x_n не будет обеспечено выполнение неравенства

$$|x_n - x_{n-1}| \leqslant \frac{1-q}{q}\varepsilon,$$

где ε — заданная предельная абсолютная погрешность корня ξ. Тогда

$$\xi = x_n \pm \varepsilon.$$

Метод итерации применяется и для решения систем уравнений. Пусть, например, дана система

$$\begin{cases} F_1(x, y) = 0, \\ F_2(x, y) = 0. \end{cases}$$

Представим ее в виде

$$\begin{cases} x = \varphi_1(x,\ y), \\ y = \varphi_2(x,\ y). \end{cases}$$

Если x_0, y_0 — приближенные значения корней системы, то по методу итерации последовательные приближения находятся по формулам

$$x_{n+1} = \varphi_1(x_n,\ y_n), \quad y_{n+1} = \varphi_2(x_n,\ y_n), \quad n \in N.$$

Для сходимости итерационного процесса на функции $\varphi_1(x, y)$, $\varphi_2(x, y)$ надо наложить некоторые ограничения. В частности, должны выполняться условия

$$\left|\frac{\partial \varphi_1}{\partial x}\right| + \left|\frac{\partial \varphi_1}{\partial y}\right| \leqslant q_1 < 1,$$

$$\left|\frac{\partial \varphi_2}{\partial x}\right| + \left|\frac{\partial \varphi_2}{\partial y}\right| \leqslant q_2 < 1.$$

Метод итерации можно использовать и для решения линейной системы уравнений, например, системы

$$\begin{cases} a_1x + b_1y + c_1z = d_1, \\ a_2x + b_2y + c_2z = d_2, \\ a_3x + b_3y + c_3z = d_3. \end{cases}$$

Предположим, что $a_1 \neq 0$, $b_2 \neq 0$, $c_3 \neq 0$ и запишем систему в виде

$$\begin{cases} x = \frac{d_1}{a_1} - \frac{b_1}{a_1}y - \frac{c_1}{a_1}z, \\ y = \frac{d_2}{b_2} - \frac{a_2}{b_2}x - \frac{c_2}{b_2}z, \\ z = \frac{d_3}{c_3} - \frac{a_3}{c_3}x - \frac{b_3}{c_3}y. \end{cases}$$

Теперь можно применить метод итерации, положив, например,

$$x_0 = \frac{d_1}{a_1}, \quad y_0 = \frac{d_2}{b_2}, \quad z_0 = \frac{d_3}{c_3}.$$

Подставив x_0, y_0, z_0 в правые части уравнений, найдем x_1, y_1 и z_1. После этого найдем x_2, y_2, z_2 и т. д.

§ 50. КОНЕЧНЫЕ РАЗНОСТИ

Пусть дана функция $y = f(x)$.

Обозначим через $\Delta x = h$ фиксированную величину приращения аргумента (шаг). Тогда выражение

$$\Delta y = \Delta f(x) = f(x + \Delta x) - f(x)$$

называется первой конечной разностью данной функции $y = f(x)$.

Аналогично определяются разности высших порядков:

$$\Delta^2 y = \Delta(\Delta y), \quad \Delta^3 y = \Delta(\Delta^2 y), \quad \ldots,$$

$$\Delta^n y = \Delta(\Delta^{n-1} y).$$

Пример. Построить конечные разности для функции $P_2(x) = x^2$, взяв шаг $\Delta x = 1$.

Решение.

$$\Delta P_2(x) = (x^2 + 1)^2 - x^2 = 2x + 1,$$

$$\Delta^2 P_2(x) = (2(x + 1) + 1) - (2x + 1) = 2,$$

$$\Delta^3 P_2(x) = 0, \ \ldots, \ \Delta^n P_2(x) = 0 \text{ при } n > 2.$$

Вообще, если $f(x) = P_n(x)$, где $P_n(x) = a_0 x^n + \ldots + a_n$ — многочлен степени n, то

$$\Delta^n P_n(x) = n!\, a_0 h^n = \text{const}, \text{ где } h \text{ — шаг},$$

$$\Delta^S P_n(x) = 0 \text{ при } S > n.$$

Символ Δ можно рассматривать как оператор, ставящий в соответствие функции $y = f(x)$ функцию $\Delta y = f(x + \Delta x) = -f(x)$, где $\Delta x = h = \text{const}$. (Не следует отождествлять Δ с оператором Лапласа, который обозначают той же буквой!)

Почти очевидны следующие свойства оператора Δ (оператора нахождения конечной разности):

$$\Delta(u \pm v) = \Delta u \pm \Delta v,$$

$$\Delta(Cu) = C\Delta u, \text{ где } C = \text{const},$$

$$\Delta^m(\Delta^n y) = \Delta^{m+n} y$$

где $m \geqslant 0$, $n \geqslant 0$ — целые, причем по определению $\Delta^0 y = y$ (т. е. сама функция $y = f(x)$ — это разность нулевого порядка). Из формулы

$$\Delta y = f(x + \Delta x) - f(x)$$

имеем

$$f(x + \Delta x) = f(x) + \Delta y = f(x) + \Delta f(x).$$

Это можно записать в символической форме так:

$$f(x + \Delta x) = (1 + \Delta) f(x).$$

Последовательно применяя это соотношение n раз, получим

$$f(x + n\Delta x) = (1 + \Delta)^n f(x) = \sum_{k=0}^{n} C_n^k \cdot \Delta^k f(x),$$

где

$$C_n^k = \frac{n(n-1) \ldots (n-k+1)}{k!}.$$

$$f(x + n\Delta x) = \sum_{k=0}^{n} C_n^k \cdot \Delta^k f(x).$$

Эта формула позволяет последовательные значения функции $f(x)$ выразить через ее конечные разности различных порядков. Например,

$$f(x+2\Delta x)=\sum_{k=0}^{2} C_2^k\cdot\Delta^k f(x)=C_2^0\cdot\Delta^0 f(x)+C_2^1\cdot\Delta f(x)+$$

$$+C_2^2\cdot\Delta^2 f(x)=f(x)+2\cdot\Delta f(x)+\Delta^2 f(x).$$

Воспользовавшись тождеством $\Delta=(1+\Delta)-1$ и применив формулу бинома Ньютона, можно показать, что

$$\Delta^n f(x)=f(x+n\Delta x)-C_n^1 f(x+(n-1)\Delta x)+$$

$$+C_n^2 f(x+(n-2)\Delta x)-\cdots+(-1)^n f(x).$$

Эта формула выражает конечную разность n-го порядка функции $f(x)$ через последовательные значения этой функции. Например, при $n=2$

$$\Delta^2 f(x)=f(x+2\Delta x)-2f(x+\Delta x)+f(x).$$

Если функция $f(x)$ имеет непрерывную производную $f^{(n)}(x)$ в $]x;\ x+n\cdot\Delta x[$, то справедлива формула

$$\Delta^n f(x)=f^{(n)}(x+\theta n\Delta x)\,\Delta x^n,$$

где $0<\theta<1$.

Тогда

$$f^{(n)}(x+\theta n\Delta x)=\frac{\Delta^n f(x)}{\Delta x^n}.$$

Переходя в последнем равенстве к пределу при $\Delta x\to 0$, получим

$$f^{(n)}(x)=\lim_{\Delta x\to 0}\frac{\Delta^n f(x)}{\Delta x^n}.$$

При достаточно малых Δx

$$f^{(n)}(x)=\frac{\Delta^n f(x)}{\Delta x^n},$$

или

$$\Delta^n f(x)\approx f^{(n)}(x)\cdot\Delta x^n.$$

Если функция $y=f(x)$ задана таблицей ее значений $y_i=f(x_i)$ для системы равноотстоящих точек x_i, где $\Delta x_i=x_{i+1}-x_i=h=$ const, то конечные разности последовательности y_i определяются по формулам:

$$\Delta y_i=y_{i+1}-y_i,$$
$$\Delta^2 y_i=\Delta(\Delta y_i)=\Delta y_{i+1}-\Delta y_i,$$
$$\cdots\cdots\cdots\cdots\cdots$$
$$\Delta^n y_i=\Delta(\Delta^{n-1}y_i)=\Delta^{n-1}y_{i+1}-\Delta^{n-1}y_i.$$

Конечные разности различных порядков обычно записывают в виде таблицы:

Таблица 15

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	...	$\Delta^n y$
x_0	y_0	Δy_0	$\Delta^2 y_0$	$\Delta^3 y_0$	.	$\Delta^n y_0$
x_1	y_1	Δy_1	$\Delta^2 y_1$	$\Delta^3 y_1$	.	$\Delta^n y_1$
x_2	y_2	Δy_2	$\Delta^2 y_2$	$\Delta^3 y_2$	.	.
x_3	y_3	Δy_3	$\Delta^2 y_3$	.	.	.
x_4	y_4	Δy_4	.	.	.	.
x_5	y_5	.	.	.	.	.
.	.	.	.	.	.	.

§ 51. ИНТЕРПОЛИРОВАНИЕ ФУНКЦИЙ

Простейшая задача интерполирования формулируется так: В $[a; b]$ заданы $n+1$ точка $a = x_0 < x_1 < x_2 < \ldots < x_n = b$ (узлы интерполяции) и значения некоторой функции $f(x)$ в этих точках $f(x_0) = y_0$, $f(x_1) = y_1$, ..., $f(x_n) = y_n$. Требуется найти многочлен $P_n(x)$ степени не выше n, который в узлах интерполяции принимает такие же значения, как и функция $f(x)$:

$$P_n(x_0) = f(x_0) = y_0, \quad P_n(x_1) = f(x_1) = y_1, \ \ldots,$$
$$P_n(x_n) = f(x_n) = y_n.$$

Многочлен $P_n(x)$ называют интерполяционным многочленом для функции $f(x)$.

Он применяется для приближенного вычисления значений данной функции $f(x)$ в точках x, отличных от узлов интерполяции. При этом, если $x \in [a; b]$, то нахождение $f(x)$ называется интерполированием, если же x берется вне $[a; b]$, нахождение $f(x)$ называется экстраполированием.

Существуют различные формулы для интерполяционного многочлена. Рассмотрим некоторые из них.

Интерполяционные формулы Ньютона. Предположим, что шаг интерполяции h постоянный. Тогда узлы интерполяции

$$x_i = x_0 + ih \quad (i = 0, 1, 2, \ldots, n).$$

Интерполяционный многочлен будем искать в виде

$$P_n(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \cdots + a_n(x - x_0) \ldots (x - x_{n-1}). \quad (5)$$

Задача заключается в нахождении коэффициентов $a_0, a_1, \ldots, a_n$.

Полагая в равенстве (5) $x = x_0$, получим, что $P_n(x_0) = a_0$. Но должно быть $P_n(x_0) = y_0$, следовательно,

$$a_0 = y_0.$$

Чтобы определить коэффициент a_1, найдем первую конечную разность:

$$\Delta P_n(x) = P_n(x+h) - P_n(x) = a_1 h + 2a_2 h(x - x_0) +$$
$$+ 3a_3 h(x - x_0)(x - x_1) + \cdots + n a_n h(x - x_0) \ldots (x - x_{x-2}).$$

Полагая здесь $x = x_0$, получим $\Delta P_n(x_0) = \Delta y_0 = a_1 h$. Тогда

$$a_1 = \frac{\Delta y_0}{1!h}.$$

Вычислив конечную разность второго порядка и положив в ней $x = x_0$, найдем

$$a_2 = \frac{\Delta^2 y_0}{2!h^2}.$$

Продолжая этот процесс, получаем, что

$$a_k = \frac{\Delta^k y_0}{k!h^k} \quad (k = 0,\ 1,\ 2,\ \ldots,\ n) \quad (0! = 1,\ \Delta^0 y_0 = y_0).$$

Тогда интерполяционный многочлен Ньютона примет вид

$$P_n(x) = y_0 + \frac{\Delta y_0}{1!h}(x - x_0) + \frac{\Delta^2 y_0}{2!h^2}(x - x_0)(x - x_1) + \cdots +$$
$$+ \frac{\Delta^n y_0}{n!h^n}(x - x_0) \ldots (x - x_{n-1}).$$

Очевидно, $P_n(x)$ — многочлен степени не выше n, причем

$$P_n(x_0) = y_0,$$
$$P_n(x_1) = y_0 + \frac{\Delta y_0}{1!h}(x_1 - x_0) = y_0 + \frac{\Delta y_0}{1!h} h = y_1,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$P_n(x_k) = y_0 + \frac{\Delta y_0}{1!h}(x_k - x_0) + \frac{\Delta^2 y_0}{2!h^2}(x_k - x_0)(x_k - x_1) + \cdots +$$
$$+ \frac{\Delta^k y_0}{k!h^k}(x_k - x_0) \ldots (x_k - x_{k-1})$$

(все остальные слагаемые равны нулю, так как имеют множитель $x_k - x_k = 0$)

$$x_k - x_0 = kh, \quad x_k - x_1 = (k-1)h, \quad \ldots, \quad x_k - x_{k-1} = h.$$

Тогда

$$P_n(x_k) = y_0 + k \cdot \Delta y_0 + \frac{k(k-1)}{2!}\Delta^2 y_0 + \cdots + \Delta^k y_0 =$$
$$= (1 + \Delta)^k y_0 = y_k \quad (k = 1,\ 2,\ \ldots,\ n).$$

Введем новую переменную:

$$u = \frac{x - x_0}{h}.$$

Теперь первая интерполяционная формула Ньютона запишется в виде

$$P_n(x) = y_0 + u \cdot \Delta y_0 + \frac{u(u-1)}{2!}\Delta^2 y_0 + \cdots +$$

$$+ \frac{u(u-1)\ldots(u-n+1)}{n!}\Delta^n y_0.$$

Переменная u в узлах интерполяции принимает значения 0, 1, 2, ..., n.

Первую интерполяционную формулу Ньютона удобно применять для интерполирования функции в окрестности точки x_0, где $u = \frac{x-x_0}{h}$ достаточно мало по абсолютной величине.

При $n = 1$ получим формулу линейного интерполирования

$$P_1(x) = y_0 + u \cdot \Delta y_0.$$

При $n = 2$ получим формулу параболического (квадратичного) интерполирования

$$P_2(x) = y_0 + u \cdot \Delta y_0 + \frac{u(u-1)}{2} \cdot \Delta^2 y_0.$$

Если таблица значений функции неограниченна, то число n в интерполяционной формуле может быть взято любым. Его практически выбирают так, чтобы разности $\Delta^n y_i$ были постоянны с заданной степенью точности.

Пример. Построить эмпирическую формулу для функции y, заданной таблицей:

Таблица 16

x	0	1	2	3	4	5
y	2,2	6,2	10,4	14,8	19,4	24,2

Решение. Составим таблицу разностей:

Таблица 17

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
0	2,2	4,0	0,2	0
1	6,2	4,2	0,2	0
2	10,4	4,4	0,2	0
3	14,8	4,6	0,2	
4	19,4	4,8		
5	24,2			

Из таблицы видно, что вторые разности постоянны, а третьи разности равны нулю. Следовательно, надо в интерполяционной формуле Ньютона взять $n = 2$. Шаг $h = 1$. Тогда

$$y = y_0 + h \cdot \Delta y_0 (x - x_0) + \frac{h^2}{2} \Delta^2 y_0 (x - x_0)(x - x_1) =$$

$$= 2{,}2 + 4x + 0{,}1 \cdot x (x - 1) = 2{,}2 + 3{,}9x + 0{,}1x^2.$$

Для интерполирования функции вблизи конца таблицы применяется вторая интерполяционная формула Ньютона:

$$P_n(x) = y_n + \frac{\Delta y_{n-1}}{1!h}(x - x_n) + \frac{\Delta^2 y_{n-2}}{2!h^2}(x - x_n)(x - x_{n-1}) + \cdots +$$

$$+ \frac{\Delta^n y_0}{n!h^n}(x - x_n) \ldots (x - x_1).$$

Если положить $\frac{x - x_n}{h} = u$, то получим

$$P_n(x) = y_n + u \cdot \Delta y_{n-1} + \frac{u(u+1)}{2!} \cdot \Delta^2 y_{n-2} + \cdots +$$

$$+ \frac{u(u+1) \ldots (u+n-1)}{n!} \Delta^n y_0.$$

Интерполяционные формулы Ньютона можно применять и для экстраполирования функций, т. е. для нахождения ее значений при значениях аргумента x, лежащих вне пределов таблицы.

При этом для $x < x_0 = a$ и достаточно близких к x_0 следует применять первую интерполяционную формулу, причем в этом случае

$$u = \frac{x - x_0}{h} < 0.$$

Если $x > x_n = b$ и близко к x_n, то следует применять вторую интерполяционную формулу. Тогда $u = \frac{x - x_n}{h} > 0$.

Как уже отмечалось, интерполяционный многочлен $P_n(x)$ в точках x, не совпадающих с узлами интерполирования $x_0, \ldots, x_n$, представляет данную функцию $y = f(x)$ лишь приближенно. Поэтому

$$y = P_n(x) + R_n(x),$$

где $R_n(x)$ — остаточный член.

Можно показать, что остаточный член первой интерполяционной формулы Ньютона имеет вид

$$R_n(x) = \frac{h^{n+1}}{(n+1)!} u(u-1) \ldots (u-n) \cdot f^{(n+1)}(\xi),$$

где $u = \frac{x - x_0}{h}$, ξ — некоторое промежуточное значение между узлами интерполирования и точкой x (в случае интерполирования ξ находится в $[x_0; x_n]$, в случае экстраполирования ξ может оказаться и вне этого промежутка).

Остаточный член второй интерполяционной формулы Ньютона

$$R_n(x) = \frac{h^{n+1}}{(n+1)!} u(u+1) \ldots (u+n) \cdot f^{(n+1)}(\xi),$$

где $u = \frac{x - x_n}{h}$, а относительно ξ справедливо сказанное выше.

Предполагая, что для функции $y = f(x)$ разности $\Delta^{n+1}y$ почти постоянны, а h — достаточно мало, и учитывая, что

$$f^{(n+1)}(x) = \lim_{h \to 0} \frac{\Delta^{n+1}y}{h^{n+1}},$$

можно считать

$$f^{(n+1)}(\xi) \approx \frac{\Delta^{n+1}y_0}{h^{n+1}}.$$

Тогда остаточный член первой интерполяционной формулы Ньютона

$$R_n(x) \approx \frac{u(u-1) \ldots (u-n)}{(n+1)!} \cdot \Delta^{n+1}y_0,$$

а остаточный член второй интерполяционной формулы Ньютона

$$R_n(x) \approx \frac{u(u+1) \ldots (u+n)}{(n+1)!} \cdot \Delta^{n+1}y_n.$$

Погрешность при вычислении приближенного значения функции $y = f(x)$ с помощью интерполяционных формул не превышает абсолютной величины соответствующего остаточного члена.

Интерполяционная формула Лагранжа. Интерполяционные формулы Ньютона пригодны лишь в случае равноотстоящих узлов интерполирования.

Для произвольно заданных узлов интерполирования применяется формула Лагранжа:

$$P_n(x) = \sum_{i=0}^{n} y_i \frac{(x - x_0) \ldots (x - x_{i-1})(x - x_{i+1}) \ldots (x - x_n)}{(x_i - x_0) \ldots (x_i - x_{i-1})(x_i - x_{i+1}) \ldots (x_i - x_n)}.$$

Если узлы интерполирования равноотстоящие ($h = \text{const}$), то интерполяционный многочлен Лагранжа совпадает с соответствующим интерполяционным многочленом Ньютона. Обозначим

$$(x - x_0)(x - x_1) \ldots (x - x_n) = \Pi_{n+1}(x).$$

Дифференцируя это равенство по x, получим

$$\Pi'_{n+1}(x) = \sum_{j=0}^{n} (x - x_0) \ldots (x - x_{j-1})(x - x_{j+1}) \ldots (x - x_n).$$

Положим здесь $x = x_i$ ($i = 0, 1, 2, \ldots, n$). Убеждаемся, что

$$\Pi'_{n+1}(x_i) = (x_i - x_0) \ldots (x_i - x_{i-1})(x_i - x_{i+1}) \ldots (x_i - x_n).$$

Тогда интерполяционную формулу Лагранжа можно кратко записать так:

$$P_n(x) = \Pi_{n+1}(x) \cdot \sum_{i=0}^{n} \frac{y_i}{\Pi'_{n+1}(x_i)(x - x_i)} =$$

$$= \sum_{i=0}^{n} \frac{\Pi_{n+1}(x) \cdot y_i}{\Pi'_{n+1}(x_i)(x - x_i)} = \sum_{i=0}^{n} L_i^{(n)}(x) \cdot y_i,$$

где $L_i^{(n)}(x) = \dfrac{\Pi_{n+1}(x)}{\Pi'_{n+1}(x_i)(x - x_i)}$ — так называемые лагранжевы коэффициенты.

Для их вычисления составим таблицу:

$$\begin{array}{llll}
\underline{x - x_0} & x_0 - x_1 & x_0 - x_2 & \ldots\ x_0 - x_n, \\
x_1 - x_0 & \underline{x - x_1} & x_1 - x_2 & \ldots\ x_1 - x_n, \\
x_2 - x_0 & x_2 - x_1 & \underline{x - x_2} & \ldots\ x_2 - x_n, \\
\ldots & \ldots & \ldots & \ldots \\
x_n - x_0 & x_n - x_1 & x_n - x_2 & \ldots\ \underline{x - x_n}.
\end{array}$$

Обозначим произведение элементов первой строки через D_0, второй — через D_1 и т. д.

Произведение элементов главной диагонали (они в таблице подчеркнуты) равно $\Pi_{n+1}(x)$. Поэтому $L_i^{(n)}(x) = \dfrac{\Pi_{n+1}(x)}{D_i}$ $(i = 0, 1, \ldots, n)$. Следовательно,

$$P_n(x) = \Pi_{n+1}(x) \cdot \sum_{i=0}^{n} \frac{y_i}{D_i}.$$

Остаточный член интерполяционной формулы Лагранжа имеет вид

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \Pi_{n+1}(x),$$

где ξ лежит внутри $[x_0; x_n]$.

Пусть

$$M_{n+1} = \max_{a \leqslant x \leqslant b} |f^{(n+1)}(x)|.$$

Тогда

$$|R_n(x)| \leqslant \frac{M_{n+1}}{(n+1)!} \cdot |\Pi_{n+1}(x)|.$$

Это неравенство дает оценку абсолютной погрешности интерполяционной формулы Лагранжа.

Если функция $f(x)$ непрерывна в $[a; b]$ и известна ее первообразная $F(x)$, то $\int_a^b f(x)\,dx = F(b) - F(a)$.

Во многих случаях первообразная $F(x)$ для функции $f(x)$ или не выражается в конечном виде через известные элементарные функции или же является очень сложной. Кроме того, часто функция $f(x)$ задается таблицей или графиком. Поэтому важное значение имеют приближённые (и в первую очередь численные) методы вычисления определенного интеграла.

Численное интегрирование заключается в вычислении определенного интеграла по ряду значений подынтегральной функции.

Обычно в этом случае данную функцию $f(x)$ ь $[a; b]$ заменяют некоторой другой функцией $\varphi(x)$ и полагают

$$\int_a^b f(x)\,dx \approx \int_a^b \varphi(x)\,dx.$$

Функция $\varphi(x)$ выбирается так, чтобы $\int_a^b \varphi(x)\,dx$ можно было вычислить сравнительно просто.

Пусть известны значения функции $y = f(x)$ в точках $a = x_0$, $x_1, x_2, \ldots, x_n = b$, причем $f(x_i) = y_i$ $(i = 0, 1, 2, \ldots, n)$.

Построим интерполяционный многочлен Лагранжа

$$P_n(x) = \sum_{i=0}^{n} \frac{\Pi_{n+1}(x)}{(x - x_i)\,\Pi'_{n+1}(x_i)} \cdot y_i,$$

где

$$\Pi_{n+1}(x) = (x - x_0)(x - x_1) \ldots (x - x_n),$$

а

$$P_n(x_i) = y_i \quad (i = 0, 1, 2, \ldots, n).$$

Тогда $$\int_a^b f(x)\,dx = \int_a^b P_n(x)\,dx + R_n,$$

где R_n — остаточный член (погрешность квадратурной формулы).

$$\int_a^b f(x)\,dx \approx \int_a^b P_n(x)\,dx = \sum_{i=0}^{n} A_i y_i,$$

где $$A_i = \int_a^b \frac{\Pi_{n+1}(x)\,dx}{(x - x_i)\,\Pi'_{n+1}(x_i)} \quad (i = 0, 1, 2, \ldots, n).$$

Коэффициенты A_i при выбранных узлах интерполяции не зависят от функции $f(x)$.

Для многочлена степени n получим точную формулу

$$\int_a^b y dx = \sum_{i=0}^{n} A_i y_i.$$

Полагая $y = x^k$ ($k = 0, 1, \ldots, n$), получаем линейную систему из $n + 1$ уравнений для определения $A_0, A_1, \ldots, A_n$:

$$I_0 = \sum_{i=0}^{n} A_i,$$

$$I_1 = \sum_{i=0}^{n} A_i x_i,$$

$$\ldots\ldots\ldots$$

$$I_n = \sum_{i=0}^{n} A_i x_i^n,$$

где

$$I_k = \int_a^b x^k dx = \frac{b^{k+1} - a^{k+1}}{k+1} \quad (k = 0, 1, 2, \ldots, n).$$

Целесообразно узлы интерполирования $x_0, x_1, x_2, \ldots, x_n$ выбирать на одинаковом расстоянии друг от друга. Тогда

$$\int_a^b y dx = \int_{x_0}^{x_n} y dx \approx \sum_{i=0}^{n} A_i y_i,$$

где коэффициенты A_i находятся по формулам

$$A_i = (x_n - x_0) H_i = (b - a) H_i,$$

$$H_i = \frac{(-1)^{n-i}}{n i! (n-i)!} \int_0^n \frac{u^{(n+1)}}{u - i} du \quad (i = 0, 1, 2, \ldots, n), \tag{6}$$

где введено обозначение:

$$u^{(n+1)} = u(u-1)(u-2) \ldots (u-n),$$

$h = \frac{b-a}{n}$ — шаг интерполирования, $u = \frac{x - x_0}{h}$.

Квадратурная формула в этом случае примет вид

$$\int_a^b y dx \approx (b - a) \sum_{i=0}^{n} H_i y_i, \tag{7}$$

где

$$y_i = f(a + ih).$$

Для контроля полезно заметить, что

$$\sum_{i=0}^{n} H_i = 1, \quad H_i = H_{n-i}.$$

Формула Симпсона. Из формулы (6) при $n = 2$ получим

$$H_0 = \frac{1}{2} \cdot \frac{1}{2} \int_0^2 (u-1)(u-2)\,du = \frac{1}{6},$$

$$H_1 = -\frac{1}{2} \cdot \frac{1}{1} \int_0^2 u(u-2)\,du = \frac{2}{3},$$

$$H_2 = \frac{1}{2} \cdot \frac{1}{2} \int_0^2 u(u-1)\,du = \frac{1}{6}.$$

Тогда

$$\int_{x_0}^{x_2} y\,dx \approx (x_2 - x_0)\left(\frac{1}{6}y_0 + \frac{2}{3}y_1 + \frac{1}{6}y_2\right) = \frac{h}{3}(y_0 + 4y_1 + y_2), \quad (8)$$

где

$$h = \frac{x_2 - x_0}{2}.$$

В данном случае кривая $y = f(x)$ в $[x_0; x_2]$ заменяется параболой $y = P_2(x)$, проходящей через точки (x_0, y_0), (x_1, y_1), (x_2, y_2) (рис. 113).

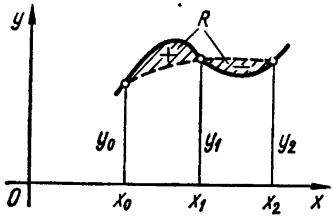

Рис. 113

Остаточный член квадратурной формулы (8) равен

$$R = -\frac{h^5}{90} y^{IV}(\xi),$$

где ξ — некоторая точка в $]x_0; x_2[$.

Разделим $[a; b]$ на четное число $n = 2m$ равных частей точками $a = x_0 < x_1 < \ldots < x_{2m} = b$.

Пусть $h = \frac{b-a}{n} = \frac{b-a}{2m}$ — шаг. Применим формулу (8) к промежуткам $[x_0; x_2]$, $[x_2; x_4]$ ..., $[x_{2m-2}; x_{2m}]$ длины $2h$. Тогда

$$\int_a^b y\,dx \approx \frac{h}{3}(y_0 + 4y_1 + y_2 + \frac{h}{3}(y_2 + 4y_3 + y_4) + \cdots +$$

$$+ \frac{h}{3}(y_{2m-2} + 4y_{2m-1} + y_{2m}).$$

В результате получим формулу Симпсона

$$\int_a^b y\,dx \approx \frac{h}{3}\left((y_0 + y_{2m}) + 4(y_1 + y_3 + \cdots + y_{2m-1}) + 2(y_2 + y_4 + \cdots + y_{2m-2})\right). \quad (9)$$

Остаточный член формулы Симпсона

$$R = -\frac{(b-a)h^4}{180} y^{IV}(\xi), \quad (10)$$

где точка ε находится в $[a; b]$.

Если задана предельная допустимая погрешность $\varepsilon > 0$, то шаг при применении формулы Симпсона находится из условия

$$h < \sqrt[4]{\frac{180\varepsilon}{(b-a)M_4}},$$

где

$$M_4 = \max |y^{IV}(x)| \text{ в } [a; b].$$

Если оценка погрешности по формуле (10) затруднительна, то рекомендуется провести по формуле (9) вычисления с шагом h и с шагом $2h$ и считать, что совпадающие в ответах десятичные знаки принадлежат к точному значению интеграла.

Если Σ_h и Σ_H — приближенные значения интеграла, вычисленные по формуле Симпсона соответственно с шагом h и с шагом $H = 2h$, то за приближенное значение интеграла целесообразно принять величину

$$\Sigma_h + \frac{\Sigma_h - \Sigma_H}{15}.$$

Удобно число делений n брать кратным четырем; тогда для вычисления Σ_H надо воспользоваться имеющимися табличными значениями функции, необходимыми для вычисления Σ_h, делая выборку значений из этой таблицы через одно значение.

Имеется и ряд других квадратурных формул (например, формулы Чебышева, Гаусса).

§ 53. ПРИБЛИЖЕННЫЕ СПОСОБЫ РЕШЕНИЯ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

Известно, что далеко не всегда удается в замкнутом виде найти решение дифференциального уравнения. Поэтому разработаны методы для их приближенного решения. Имеется большое количество таких методов. Здесь будут рассмотрены лишь некоторые из них.

Метод Эйлера. Найдем приближенно решение уравнения $y' = f(x, y)$ при начальном условии $y(x_0) = y_0$ в точках $x_1 = x_0 + h, \ldots, x_n = x_0 + nh$, где шаг h — постоянный.

Обозначим $y(x_k) = y_k$. Учитывая, что

$$y'(x_k) \approx \frac{\Delta y_k}{h} = \frac{y_{k+1} - y_k}{h},$$

заданное уравнение запишем так:

$$\frac{y_{k+1}-y_k}{h}=f(x_k,\ y_k).$$

Отсюда

$$y_{k+1}=y_k+hf(x_k,\ y_k).$$

Полагая в этой формуле $k=0, 1, 2, \ldots$, находим последовательно величины $y_1, y_2, y_3, \ldots$.

По этим данным построим ломаную Эйлера, которая приближенно заменяет интегральную кривую, проходящую через точку $M_0\ (x_0,\ y_0)$ (рис. 114).

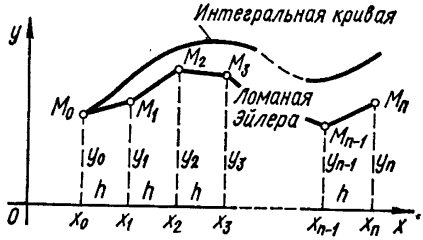

Рис. 114

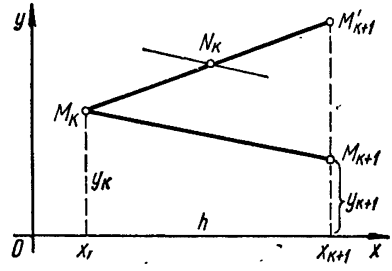

Рис. 115

Геометрически метод Эйлера означает, что через точку $M_0\ (x_0,\ y_0)$ проводим не интегральную кривую, а отрезок касательной к ней, который и принимаем приближенно за интегральную кривую на участке от x_0 до $x_1=x_0+h$. После этого за исходную точку берем $M_1\ (x_1,\ y_1)$ и поступаем так же, как и до этого.

Метод Эйлера прост, но точность его невелика. Если $[x_0;\ x]$, где ищется решение, разбит на n частей $h=\frac{x-x_0}{n}$, то погрешность при вычислении ординат имеет порядок $\frac{1}{n}$. Это означает, что для повышения точности в 10 раз (получения одного дополнительного десятичного знака) требуется увеличить число точек деления в десять раз. Последнее значительно увеличит объем вычислений.

Метод Рунге — Кутта. Пусть y_k — приближенное значение решения при $x=x_k$ уже построено.

Найдем $f_k=f(x_k,\ y_k)$, $\alpha_k=f\left(x_k+\frac{h}{2},\ y_k+\frac{f_kh}{2}\right)$ и за приближенное значение решения при $x=x_{k+1}=x_k+h$ примем $y_{k+1}=y_k+a_kh$.

Геометрически это означает, что для построения звена M_kM_{k+1} ломаной, аппроксимирующей интегральную кривую, сначала проводим через точку M_k отрезок $M_kM'_{k+1}$, совпадающий с направлением f_k поля в точке M_k (как это делалось в методе Эйлера). Вычисляем координаты точки N_k — середины этого отрезка: $x_k+\frac{h}{2},\ y_k+\frac{hf_k}{2}$ и находим величину $\alpha_k=f\left(x_k+\frac{h}{2},\ y_k+\frac{hf_k}{2}\right)-$

направление поля в точке N_k. Через точку M_k проводим отрезок M_kM_{k+1} параллельно направлению поля в точке N_k. Ординату точки M_{k+1} и принимаем за y_{k+1}.

Этот метод будет точнее метода Эйлера, так как он учитывает поворот поля на интервале $]x_k;\ x_{k+1}[$ (рис. 115).

Еще более точный результат получим, если вычисления проводить по схеме:

$$f_k = f(x_k,\ y_k),\quad \alpha_k = f\left(x_k + \frac{h}{2},\ y_k + \frac{f_k h}{2}\right),$$

$$\beta_k = f\left(x_k + \frac{h}{2},\ y_k + \frac{\alpha_k h}{2}\right),\quad \gamma_k = f(x_k + h,\ y_k + \beta_k h),$$

$$y_{k+1} = y_k + \frac{1}{6}(f_k + 2\alpha_k + 2\beta_k + \gamma_k)h.$$

За исходную точку, как и раньше, берем точку $M_0(x_0, y_0)$. Выбрав шаг h, вычислим:

$$f_0 = f(x_0,\ y_0),\quad \alpha_0 = f\left(x_0 + \frac{h}{2},\ y_0 + \frac{f_0 h}{2}\right),$$

$$\beta_0 = f\left(x_0 + \frac{h}{2},\ y_0 + \frac{\alpha_0 h}{2}\right),\quad \gamma_0 = f(x_0 + h,\ y_0 + \beta_0 h).$$

Находим

$$y_1 = y_0 + \frac{1}{6}(f_0 + 2\alpha_0 + 2\beta_0 + \gamma_0)h.$$

Получили точку $M_1(x_1, y_1)$. Аналогично находим f_1, α_1, β_1, γ_1, вычисляем y_2 и т. д.

Суммарная погрешность этого метода имеет порядок $\frac{1}{n^4}$.

§ 54. ПРИБЛИЖЕННЫЕ МЕТОДЫ РАСЧЕТА ПЕРЕХОДНЫХ ПРОЦЕССОВ В НЕЛИНЕЙНЫХ ЭЛЕКТРИЧЕСКИХ ЦЕПЯХ

Известно, что переходные процессы в нелинейных электрических цепях описываются нелинейными дифференциальными уравнениями, для нахождения решений которых не существует общего аналитического метода. Поэтому при исследовании переходных процессов в нелинейных цепях часто применяются приближенные методы расчета.

Наиболее широкое применение получили следующие пять методов:

1) Метод условной линеаризации характеристики. В этом методе характеристика нелинейного элемента цепи условно заменяется прямой линией. Метод имеет малую точность и используется лишь при ориентировочных расчетах.

2) Метод аналитической аппроксимации нелинейной характеристики. Здесь нелинейная характеристика приближенно заменяется некоторой функцией так, чтобы полученное при этом нелинейное

дифференциальное уравнение удалось решить. Эффективность метода зависит от удачного выбора формул для аппроксимации характеристики нелинейного элемента цепи.

3) Метод кусочно-линейной аппроксимации характеристики. В этом методе нелинейная характеристика заменяется ломаной линией. Для каждого звена этой ломаной получают линейное дифференциальное уравнение с постоянными коэффициентами. Решения этих уравнений «стыкуются» («припасовываются») в вершинах ломаной путем соответствующего выбора постоянных интегрирования из условий непрерывности токов в ветвях цепи с индуктивностью и напряжений на конденсаторах.

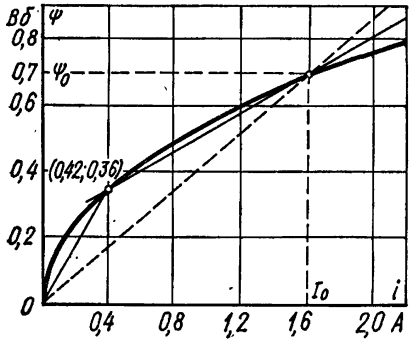

Рис. 116

4) Метод последовательных интервалов. Здесь дифференциальное уравнение цепи заменяется обычным алгебраическим уравнением, содержащим конечные приращения исследуемой величины за достаточно малые интервалы времени, на которые делится весь интервал времени, где рассматривается переходный процесс. Это метод численного интегрирования. Он простой, но довольно трудоемкий.

5) Метод графического интегрирования. Сущность этого метода заключается в преобразовании нелинейного дифференциального уравнения цепи к виду, для которого можно построить такую кривую, чтобы площадь, ограниченная ею и осью ординат, была пропорциональна времени, соответствующему данному значению исследуемой переходной величины. Этим методом достаточно просто и точно решаются дифференциальные уравнения первого порядка, допускающие разделение переменных.

Пример. Катушка со стальным сердечником, подключенная к постоянному напряжению $U = 100$ В, замыкается накоротко. Зависимость потокосцепления от тока приближенно задана уравнением $i = 3{,}27\Psi^2$, где i— в амперах, Ψ — в веберах (рис. 116). Сопротивление катушки $r = 62{,}4$ Ом.

Найти закон изменения тока переходного процесса в цепи:

а) аналитически;

б) методом условной линеаризации, решая уравнение относительно тока;

в) методом кусочно-линейной аппроксимации, заменяя нелинейную характеристику $i(\Psi)$ двумя прямыми (рис. 116);

г) методом последовательных интервалов;

д) методом графического интегрирования.

Р е ш е н и е. а) Аналитическое решение задачи ищем, исходя из заданной аппроксимации характеристики формулой $i = 3{,}27\Psi^2$.

Переходный процесс в рассматриваемой цепи описывается уравнением

$$ri + \frac{d\Psi}{dt} = 0,$$

которое в данном случае будет иметь вид

$$62{,}4 \cdot 3{,}27\Psi^2 + \frac{d\Psi}{dt} = 0,$$

или

$$204\Psi^2 + \frac{d\Psi}{dt} = 0.$$

Выполним в этом уравнении разделение переменных и решим его относительно времени t:

$$-204\Psi^2 dt = d\Psi, \quad dt = -0{,}49 \cdot 10^{-2} \frac{d\Psi}{\Psi^2},$$

$$t = -0{,}49 \cdot 10^{-2} \int_{\Psi_0}^{\Psi} \frac{d\Psi}{\Psi^2} = \frac{0{,}49 \cdot 10^{-2}}{\Psi} \Big|_{\Psi_0}^{\Psi} = 0{,}49 \cdot 10^{-2} \left(\frac{1}{\Psi} - \frac{1}{\Psi_0} \right),$$

где $\Psi_0 = 0{,}7$ Вб — установившееся значение потокосцепления до начала переходного процесса (при $t = 0$), определенное по кривой $\Psi(i)$ по установившемуся значению тока в начальный момент переходного процесса $i(0) = I_0 = \frac{U}{r} = \frac{100}{62{,}4} = 1{,}6$ А.

Поэтому

$$t = \frac{0{,}49 \cdot 10^{-2}}{\Psi} - 0{,}7 \cdot 10^{-2} \text{ с},$$

откуда получим

$$\Psi = \frac{0{,}49 \cdot 10^{-2}}{t + 0{,}7 \cdot 10^{-2}} \text{ Вб}.$$

Подставив найденное значение Ψ в выражение для i, найдем

$$i = 3{,}27 \left(\frac{0{,}49 \cdot 10^{-2}}{t + 0{,}7 \cdot 10^{-2}} \right)^2 \text{ А}.$$

б) Для решения задачи методом условной линеаризации относительно тока через точку, соответствующую начальным условиям, на кривой $\Psi(i)$ проводим прямую $\Psi = L_0 i$ (рис. 116), где

$$L_0 = \frac{\Psi_0}{I_0} = \frac{0{,}7}{1{,}6} = 0{,}437 \text{ Г}.$$

Теперь уравнение цепи $ri + \frac{d\Psi}{dt} = 0$ можно записать в виде

$$ri + L_0 \frac{di}{dt} = 0.$$

Получившееся линейное дифференциальное уравнение имеет решение

$$i = A_0 e^{-\frac{r}{L_0} t},$$

где $A_0 = I_0 = 1{,}6$ А, так как при $t = 0$ $i(0) = I_0 = A_0$.

$$\frac{r}{L_0} = \frac{62{,}4}{0{,}437} = \frac{100}{0{,}7}\,\frac{1}{\text{с}}.$$

Таким образом, получим $i = 1{,}6e^{-\frac{100}{0{,}7}t}$ А.

в) В методе кусочно-линейной аппроксимации характеристику $\Psi(i)$ разбиваем на два линейных участка (рис. 116).

По условию задачи на первом участке

$$\Psi = \Psi_0 \div \Psi_1 = 0{,}36 \text{ Вб}; \quad i = I_0 \div I_1 = 0{,}424 \text{ А};$$

$$t = 0 \div t, \text{ где } \Psi_0 = 0{,}7 \text{ Вб}, \ I_0 = 1{,}6 \text{ А}.$$

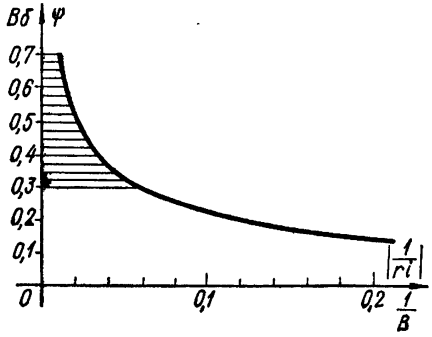

Рис. 117

При этом уравнение цепи запишется в виде

$$ri + L_1 \frac{di}{dt} = 0,$$

где $L_1 = \frac{\Delta\Psi}{\Delta I} = \frac{\Psi_0 - \Psi_1}{I_0 - I_1} = \frac{0{,}7 - 0{,}36}{1{,}6 - 0{,}424} = \frac{0{,}34}{1{,}176} = 0{,}289$ Г, и будет иметь решение

$$i = A_1 e^{-\frac{r}{L_1}t} = 1{,}6e^{-\frac{62{,}4}{0{,}289}t} = 1{,}6e^{-\frac{1000}{4{,}615}t} \text{ А},$$

где $A_1 = 1{,}6$ А, так как $i(0) = I_0 = A_1 = 1{,}6$ А, $\tau_1 = \frac{L_1}{r} = \frac{0{,}289}{62{,}4} = 0{,}0046$ с = = 4,6 мс — постоянная времени цепи для этого участка.

Найдем момент времени t_1, соответствующий значению тока $i(t_1) = I_1 = 0{,}424$ А:

$$i(t_1) = 0{,}424 = 1{,}6e^{-\frac{1000t_1}{4{,}615}}; \quad \frac{0{,}424}{1{,}6} = e^{-\frac{1000}{4{,}615}t_1},$$

$$e^{\frac{1000}{4{,}615}t_1} = \frac{1{,}6}{0{,}424} = 3{,}77, \quad \frac{1000}{4{,}615}t_1 = \ln 3{,}77,$$

$$t_1 = 4{,}615 \cdot 10^{-3} \ln 3{,}77 = 6{,}13 \cdot 10^{-3} \text{ с}.$$

Таким образом, первый участок ломаной характеристики занимает во времени интервал $t = 0 \div t_1 = 6{,}13$ мс.

На втором участке по условию (рис. 116):

$$\Psi = \Psi_1 \div 0, \quad i = I_1 \div 0; \quad t = t_1 \div \infty.$$

Здесь $L_2 = \frac{\Psi_1}{I_1} = \frac{0{,}36}{0{,}424} = 0{,}85$ Г, и уравнение цепи будет иметь вид

$$ri + L_2 \frac{di}{dt} = 0.$$

В решении этого уравнения

$$i = A_2 e^{-\frac{r}{L_2}(t - t_1)} = A_2 e^{-\frac{62{,}4}{0{,}85}(t - t_1)} = A_2 e^{-\frac{100}{1{,}36}(t - t_1)},$$

$\tau_2 = \frac{L_2}{r} = \frac{0{,}85}{62{,}4} = 0{,}0136$ с — постоянная времени цепи для данного участка, а A_2 должна быть выбрана путем «припасовывания» решений для обоих участков в момент времени $t_1 = 6{,}13$ мс.

Так как ток в индуктивности не может меняться скачком, то при $t = t_1$ будем иметь: $i(t_1) = 0{,}424 = A_2 e^0 = A_2$.

Поэтому

$$i(t) = 0{,}424 e^{-\frac{100}{1{,}36}(t-t_1)} = 0{,}424 e^{-\frac{100}{1{,}36}t_1} \cdot e^{-\frac{100}{1{,}36}t} =$$

$$= 0{,}424 e^{-\frac{100}{1{,}36}\cdot 6{,}13\cdot 10^{-3}} \cdot e^{-\frac{100}{1{,}36}t} = 0{,}665 e^{-\frac{100}{1{,}36}t} \text{ А}.$$

г) Для решения задачи методом последовательных интервалов записываем уравнение цепи $ri + \frac{d\Psi}{dt} = 0$ в конечных приращениях:

$$ri_k + \frac{\Delta\Psi_{k+1}}{\Delta t} = 0,$$

где $\Delta\Psi_{k+1} = \Psi_{k+1} - \Psi_k$ — приращение потокосцепления на $(k+1)$ интервале за промежуток времени Δt.

В результате получим следующие расчетные уравнения:

$$\Delta\Psi_{k+1} = -ri_k\Delta t, \quad \Psi_{k+1} = \Psi_k + \Delta\Psi_{k+1},$$

где i_{k+1} находим по графику $\Psi(i)$ или, в нашем случае, по формуле $i_{k+1} = 3{,}27\Psi^2_{k+1}$.

Для промежутка времени от 0 до 10 мс возьмем $\Delta t = 1$ мс.

Расчет величин Ψ_{k+1} и i_{k+1} сведен в таблицу:

Таблица 18

t	$\lvert ri_k \rvert$	$\lvert \Delta\Psi_{k+1} \rvert$	Ψ_{k+1}	i_{k+1}
мс	В	Вб	Вб	А
0	0	0	0,7000	1,600
1	99,84	0,0998	0,6002	1,178
2	73,51	0,0735	0,5267	0,907
3	56,60	0,0566	0,4701	0,723
4	45,09	0,0451	0,4250	0,665
5	41,50	0,0415	0,3835	0,471
6	29,39	0,0294	0,3541	0,410
7	25,58	0,0256	0,3285	0,353
8	22,03	0,0220	0,3065	0,307
9	19,16	0,0192	0,2873	0,270
10	16,85	0,0168	0,2705	0,289

д) В методе графического интегрирования уравнение цепи решаем относительно времени

$$ri + \frac{d\Psi}{dt} = 0, \quad t = -\int_{\Psi_0}^{\Psi} \frac{d\Psi}{ri},$$

где t пропорционально заштрихованной площади (рис. 117).

Таблица 19

Ψ	i	ri_k	$\lvert a_k \rvert$	$\frac{\lvert a_k \rvert + \lvert a_{k+1} \rvert}{2}$	Δt_{k+1}	t
Вб	А	В	1/В	1/В	мс	мс
0,70	1,6000	99,840	0,01002	0	0	0
0,65	1,3816	86,210	0,01160	0,01081	0,5405	0,540
0,60	1,1772	73,457	0,01361	0,01260	0,6300	1,170
0,55	0,9892	61,724	0,01620	0,01490	0,7450	1,915
0,50	0,8175	51,012	0,01960	0,01790	0,8950	2,810
0,45	0,6622	41,319	0,02420	0,02190	1,0950	3,905
0,40	0,5232	32,647	0,03100	0,02760	1,3800	5,285
0,35	0,4006	24,996	0,04000	0,03550	1,7750	7,060
0,30	0,2943	18,364	0,05446	0,04723	2,3615	9,421

Вычисление t выполним графоаналитически, для чего зададимся приращением $\Delta\Psi$ и введем обозначение $a_k = -\frac{1}{ri_k}$.

Тогда расчетными формулами будут равенства:

$$t_{k+1} = t_k + \Delta t_{k+1},$$

$$\Delta t_{k+1} = \frac{a_k + a_{k+1}}{2}\Delta\Psi$$

и

$$a_{k+1} = -\frac{1}{ri_{k+1}},$$

где i_{k+1} находим по графику $\Psi(i)$ или по формуле

$$i_{k+1} = 3{,}27\Psi^2_{k+1}.$$

Расчет величин t для интервала времени от 0 до 10 мс сведен в таблицу, где взято $\Delta\Psi = -0{,}05$ Вб.

В практике расчета переходных процессов в автоматических системах часто применяется приближенный графоаналитический метод Д. Б. Башкирова[1]. Этот метод позволяет получить с одинаковой простотой кривую переходного процесса как для линейных, так и для нелинейных систем, одномерных и многомерных при любом виде возмущающего воздействия.

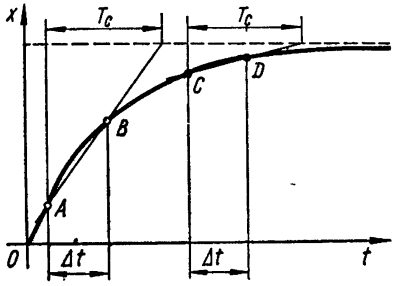

Рис. 118

Метод Башкирова основывается на принципе разложения, согласно которому большинство уравнений высокого порядка, как линейных, так и нелинейных, встречающихся при решении задач автоматического регулирования, могут быть разложены на уравнения первого порядка. Ре-

[1] См.: Д. Б. Башкиров. Графоаналитический метод построения переходных процессов в системах автоматического регулирования. Л., изд-во ЛКВВИА, 1952.

шениями последних являются экспоненты, построение которых не вызывает затруднений, так как они обладают свойством равномерности затухания. Это свойство заключается в том, что проекции секущих AB и CD (рис. 118), проведенных через две пары точек экспоненты, отстоящих друг от друга на равные промежутки времени Δt, на ось, к которой асимптотически приближается экспонента, имеют одинаковую длину T_c.

Для получения алгоритма графического построения кривой переходного процесса на основе использования указанного свойства экспоненты дифференциальное уравнение первого порядка

$$T\frac{dx}{dt} + x = f(t)$$

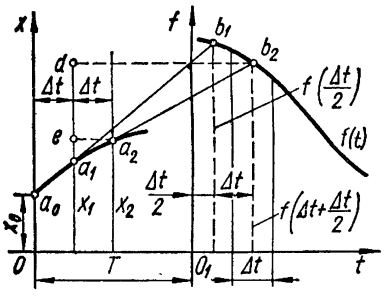

Рис. 119

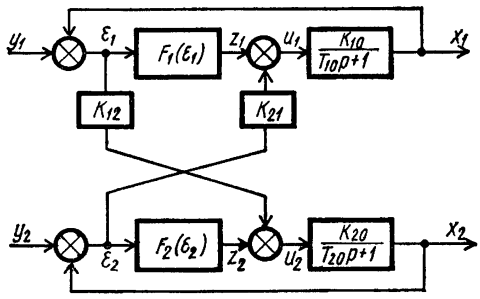

Рис. 120

заменяют приближенным уравнением в конечных приращениях:

$$\frac{\Delta x}{\Delta t} = \frac{f\left(t+\frac{\Delta t}{2}\right) - x}{T_c},$$

где

$$T_c = T + \frac{\Delta t}{2}.$$

Построение кривой переходного процесса, описываемого этим уравнением, выполняют следующим образом.

На графике (рис. 119) располагают систему исходных координат x, t и на расстоянии T от начала координат O размещают вторую систему координат f, t. На оси ординат Ox откладывают начальное значение $x(0) = x_0$ (точка a_0), а на плоскости fO_1t вычерчивают график функции $f(t)$. От точек O и O_1 проводят вертикальные прямые на расстоянии Δt друг от друга. На кривой $f(t)$ отмечают точки b_1, b_2 посередине интервалов Δt. Затем, соединяя точки b_1 и a_0, находят точку a_1, принадлежащую экспоненте. Для получения очередной точки искомой кривой переходного процесса соединяют прямой точки b_2 и a_1 и на пересечении этой прямой с ординатой, удаленной от точки O на $2\Delta t$, находят точку a_2. Так поступают и на последующих этапах построения. В результате получается полный график

кривой, проходящей через точки a_0, a_1, a_2, ... Построение будет тем точнее, чем меньше выбран шаг Δt.

Доказательство справедливости алгоритма рассмотренного построения нетрудно получить из подобия треугольников a_1ea_2 и a_1db_2 (рис. 119).

Пример. Используя метод Башкирова, построить кривые переходных процессов в двумерной нелинейной автоматической системе [1] (рис. 120).

Р е ш е н и е. Уравнение системы, соответствующей структурной схеме (рис. 120), может быть приведено к виду

$$(T_{10}p + 1)\,\varepsilon_1 + k_{10}F_1(\varepsilon_1) + k_{10}k_{21}\varepsilon_2 = (T_{10}p + 1)\,y_1,$$
$$(T_{20}p + 1)\,\varepsilon_2 + k_{20}F_2(\varepsilon_2) + k_{20}k_{12}\varepsilon_1 = (T_{20}p + 1)\,y_2.$$

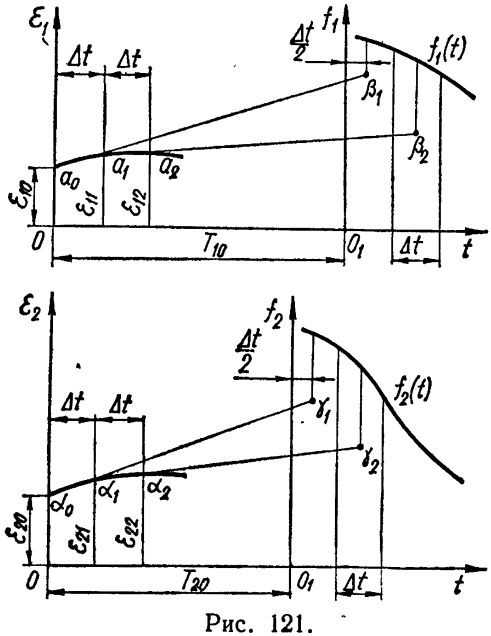

Рис. 121.

Зная задания y_1 и y_2 и их производные $\dot{y}_1$ и $\dot{y}_2$, определим правые части уравнений и обозначим их для краткости f_1 и f_2.

Решим первое и второе уравнения соответственно относительно ε_1 и ε_2:

$$(T_{10}p + 1)\,\varepsilon_1 = f_1 - k_{10}k_{21}\varepsilon_2 - k_{10}F_1(\varepsilon_1),$$
$$(T_{20}p + 1)\,\varepsilon_2 = f_2 - k_{20}k_{12}\varepsilon_1 - k_{20}F_2(\varepsilon_2).$$

Пусть к этим уравнениям заданы начальные условия ε_{10} и ε_{20} и нам известны нелинейные характеристики $F_1(\varepsilon_1)$ и $F_2(\varepsilon_2)$. Тогда правило построения решений для ε_1 и ε_2 (для обоих контуров системы) будет аналогичным правилу, показанному на рис. 119. Отличие будет состоять только в получении точек β_1, β_2 и γ_1, γ_2 (рис. 121). Так, в частности, чтобы найти точку β_1, надо из координаты $f_1\left(\frac{\Delta t}{2}\right)$ вычесть величину $(k_{10}k_{21}\varepsilon_{20} + k_{10}F_1(\varepsilon_{10}))$, а для нахождения точки γ_1 достаточно из координаты $f_2\left(\frac{\Delta t}{2}\right)$ вычесть сумму $(k_{20}k_{10}\varepsilon_{10} + k_{20}F_2(\varepsilon_{20}))$.

Поскольку контуры двумерной системы работают в едином времени, то на графиках следует иметь одинаковые значения Δt, чтобы вновь получаемые $\varepsilon_1(\Delta t)$ и $\varepsilon_2(\Delta t)$ использовать на следующих этапах решения.

Продолжая построения для ряда этапов, получим полную картину переходных процессов в первом и втором контурах системы.

Пересчет на x_1 и x_2 осуществляется просто, поскольку x_1, x_2 и ε_1, ε_2 связаны между собой с помощью известного задания y_1 и y_2.

Спецификой решения задачи в многомерной системе является неоднократное обращение от одной нелинейности к другой и от одной координаты к другой.

[1] См.: П. И. Ч и н а е в. Методы анализа и синтеза многомерных автоматических систем. К., Техніка, 1969, с. 245—247.

СПИСОК ЛИТЕРАТУРЫ

Демидович Б. П., Марон И. А. Основы вычислительной математики. М., Наука, 1966.

Еремин И. И., Астафьев Н. Н. Введение в теорию линейного и выпуклого программирования. М., Наука, 1976.

Заславский Ю. Л. Сборник задач по линейному программированию. М., Наука, 1969.

Конторович М. И. Операционное исчисление и процессы в электрических цепях. М., Советское радио, 1975.

Краснов М. Л., Киселев А. И., Макаренко Г. И. Функции комплексного переменного. Операционное исчисление. Теория устойчивости. М., Наука, 1971.

Краснов М. Л., Макаренко Г. И., Киселев А. И. Вариационное исчисление. М., Наука, 1973.

Мальцев А. И. Основы линейной алгебры. М., Наука, 1975.

Проскуряков И. В. Сборник задач по линейной алгебре. М., Наука, 1974.

Романовский П. И. Ряды Фурье. Теория поля. Аналитические и специальные функции. Преобразование Лапласа. М., ГИТТЛ, 1967.

Тихонов А. Н., Самарский А. А. Уравнения математической физики. М., Наука, 1966.

Цлаф Л. Я. Вариационное исчисление и интегральные уравнения (Справочное руководство). М., Наука, 1966.

Эльсгольц Л. Э. Дифференциальные уравнения и вариационное исчисление. М., Наука, 1969.

Аксиомы линейного векторного пространства 209

Базис векторного пространства 212
Бесконечномерные пространства 212

Вариация функционала вторая 245
— — первая 244
Вектор оператора собственный 224
Векторы линейно-зависимые 210
— линейно-независимые 210
Вихрь векторного поля 163
Волна плоская 87
— стоячая 87
— сферическая 87, 97, 98
— цилиндрическая 87
Вычет аналитической функции 40
— логарифмический 45, 46

Градиент 147

«Делители нуля» 204
Дельта-функция 96, 97
Дивергенция вектора 157, 158

Задача Больца 265
— вариационная простейшая 247
— вариационного исчисления основная 242
— изопериметрическая 240, 261
— линейного программирования 298, 299
— транспортная 293
Задачи вариационные с подвижными границами 265
— — с угловыми точками 269
— — с ограничениями 270
— об оптимальном управлении 270
Значение оператора собственное 224

Изломы экстремали 269, 270
Изображение 104, 219
— единичной функции 106
— импульсной функции 121
— интеграла от функции 109
— кусочно-непрерывных и кусочно-гладких функций 122
— периодических функций 119
— производной 107
— производной любого порядка 108
— ступенчатой функции 126
— функции по Карсону 104
Интеграл Дюамеля 120
— Лапласа 104

Косинус-преобразование функции 59
Кривая жорданова 12
Кривые смежные 242

Матрица взаимная 205
— вырожденная 195
— диагональная 195
— единичная 195
— кососимметрическая 197
— линейного преобразования 193
— невырожденная 195
— нулевого оператора 221
— оператора в базисе 220
— оператора подобия 221
— ортогональная 207
— присоединения 205
— расширения 228
— системы 228
— треугольная 195
— унитарная 207
— характеристическая 238
— эрмитова 207
— эрмитово-сопряженная 208
Матрица-столбец 194
Матрица-строка 194
Матрицы конформные 203
Методы вариационного исчисления прямые 283
Многоугольник допустимых значений 310
Многочлен оператора характеристический 225

Область 11
— замкнутая 11

— односвязная 12
Оболочки линейные 217
Обращение матриц 205
Оператор Гамильтона 170
— единичный 219
— Лапласа 173
— линейный 219
— подобия 219
Оригинал 103, 219
Отображение конформное 49

План задачи линейного программирования 313
— опорный 313
Плотность спектральная 65
Подпространство линейное 216
Поле источника 180
— потенциальное 174, 176
— соленоидальное 159, 182
— экстремалей 276
Потенциал векторный 182
— поля скалярный 174
Поток вектора 152, 153
Преобразование обратное линейное 206
— — Лапласа 143
— — Фурье 58
Преобразования матрицы элементарные 199
Принцип оптимальности 324
Пространство векторное комплексное 213
— — линейное 208
Псевдоэкстремаль 270

Размерность векторного пространства 211
Ранг матрицы 197
— системы векторов 214
— — уравнений 229
Решение допустимое 300
— оптимальное 300
Ряд Лорана 35

Свертка функций 115
Симметрическая матрица 197
Синус-преобразование функции 59
Система решений фундаментальная 233
— — — нормальная 233

Теорема запаздывания 113
— Коши 21
— — для многосвязной области 23
— Кронекера — Капелли 229
— Лиувилля 25
— Морера 25
— о вычетах основная 42
— о дифференцировании изображения 110
— о свертке 115
— об интегрировании изображения 111
— Остроградского — Гаусса 156, 161
— подобия 112
— разложения вторая 117
— разложения первая 116
— Руше 48
— смещения 114
— Стокса 167
Точка кратная 12
— особая изолированная 38
— правильная 37
— существенно особая 39
Точки особые 38

Уравнение волновое 76, 97
Условие Вейерштрасса 281
— Лежандра 250
Условия Вейерштрасса — Эрдмана 270
— д'Аламбера — Эйлера 15
— трансверсальности 267

Формула обращения Римана — Мелина 143
— Симпсона 353
— Фурье 54
Формы канонические уравнений 74, 75
Функционал линейный 243, 314
— непрерывный 243
Функция аналитическая 17
— Бесселя 92, 93, 95
— Беселя второго рода 93
— Вейерштрасса 281
— волновая 87
— гамма (определение) 90
— голоморфная 39
— единичная 104
— импульсная 96, 97
— комплексной переменной 12
— мероморфная 39
— Неймана (Вебера) 93
— однолистная 50
— Понтрягина 288, 289
— разрушения 297
— регулярная 17
— цели 304

Характеристика спектральная 65

Экстремаль 246
Экстремум функционала 245, 246, 282
— — сильный 243
— — слабый 243

Петр Иванович Чинаев,
Николай Александрович Минин,
Аскольд Юрьевич Перевозников,
Александр Александрович Черенков

ВЫСШАЯ МАТЕМАТИКА

Специальные главы

Под общ. ред. проф. П. И. Чинаева

Редактор Л. П. Онищенко
Литредактор А. П. Ковальчук
Переплет художника Г. М. Балюна
Художественный редактор Е. В. Чурий.
Технический редактор А. И. Омоховская
Корректор Н. В. Волкова

Информ. бланк № 5938

Сдано в набор 27.06.80. Подп. в печать 26.01.81. Формат 60×90¹/₁₆. Бумага типогр. № 3. Лит. гарн. Выс. печать. 23 печ. л. 23,31 кр.-отт. 23,02 уч.-изд. л. Тираж 15 000 экз. Изд. № 5107. Заказ № 673. Цена 90 к.

Головное издательство издательского объединения
«Вища школа», 252054, Киев-54, ул. Гоголевская, 7.

Напечатано с матриц Головного предприятия республиканского производственного объединения «Поліграфкнига» Госкомиздата УССР, г. Киев, Довженко, 3 на Белоцерковской книжной фабрике, 256400, г. Белая Церковь, ул. Карла Маркса, 4.